

IDENTIFICATION AND CHARACTERIZATION OF GENES  
INVOLVED IN MUTAGENESIS IN ESCHERICHIA COLI

by

STEPHEN JOSEPH ELLEDGE

B. S., University of Illinois (1978)

SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE  
DEGREE OF

DOCTOR OF PHILOSOPHY

at the  
MASSACHUSETTS INSTITUTE OF TECHNOLOGY  
January, 1983

Signature of Author

1

Department of Biology

Certified by

Graham C. Walker  
Thesis Supervisor

Accepted by

Chairman, Biology Department Graduate Committee

Archives

MASSACHUSETTS INSTITUTE  
OF TECHNOLOGY

FEB 9 1983

LIBRARIES

To my parents Joe and Sarah  
and to Charles "Ned" Holt

### Acknowledgements

I would like to express my sincere thanks to Graham Walker, my advisor, for his constant encouragement and support especially in the early stages of this work. I would like to thank him for the many discussions, ideas, criticisms and especially, for teaching me how to write papers and finally for his patience, honesty and expert scientific judgment.

I would like to thank the many people who have shared their scientific insight with me through innumerable discussions - Graham Walker, Keith Backman, Bill Shanabruch, Pam Langer, Cindy Kenyon, Russ Mauer, Maury Fox, Mike Lichten, Maria Jasin, Mark Rose, Susanna Lewis, Scott Putney, David Shortle, Iva Greenwald, Paul Sternberg, Chip Ferguson, Matt Sachs, Steve Sandler, Dennis Ballinger, Birdi Farrell, Linda Haigh, Lorraine Marsh, Teresa Keng, Rich Freedman, Judy Swan, Aaron Mitchell, Barry Wanner, Ed Loechler, Carl Falco, Karen Perry, Steve Winans, Jon Geiger, Patty Pang, Judy Kreuger, Rick Kreuger, Athena Andreadis, Lynne Vales, Pete LeMotte, Karen Overbye, Pat Foster, Eric Eisenstadt, Roy Smith, Margaret Duncan, Marilyn Lichtman and Shoshana Klein.

I would like to thank my many friends who have made my stay in Cambridge a most enjoyable one - Amy Behrman, Cindy Gray, Jon Yerkes, Art Spivack, Maria Jasin, Anita Honkanen, Lalitha Vaidyanathan, Nancy Andrews, Scott Putney, Birdi Farrell, Jenny Green, Jon Geiger, Roy Smith, Pam Langer, Paul Sternberg, Chip Ferguson, Diane Martin, Matt Sachs, Judy Kreuger, Cheryl Ann Hanna, Judy Swan, John Mulligan, Tom Weisman, Lisa Warenski and Andrea Siani. My special thanks goes to Amy Behrman, Susanna Lewis, Bill Shanabruch and Barbara Mitchell whose sincerity, good-natures, support and love have given me balance, great pleasure and occasional glimpses of sanity - not to mention a few hangovers.

I would like to thank everyone associated with the Walker lab over the years and especially to Linda Withers for all of her good natured help with the manuscripts.

I would like to thank David Wishart my college roommate and life long friend for pointing out molecular biology to me by saying "Hey, Steve, this is really neat, DNA makes RNA which makes proteins..." I would also like to thank Mike Lichten for his comment "Hey, kid, didn't anyone ever tell you that the biggest experiment you do here is the one you do on yourself" and Bill Shanabruch for reminding me that getting an education at M.I.T. is like trying to get a drink of water from a fire hydrant.

Most importantly, I would like to thank my parents Joe and Sarah and my friend Joyce for their continued love and support throughout my graduate career. They have made all the difference. I would like to thank my father whose personal strength, insight into human nature, and genuine concern for people has provided me with a standard of personal excellence which shall guide me for the rest of my life. I would also like to thank my grandmother Teresa and my sisters Billie, Barbara and especially Brenda for their support and just for being themselves.

Lastly, I would like to thank Barbara Doran for the time and care she put into preparing this thesis.

### Abstract

In *Escherichia coli*, mutagenesis by agents such as ultraviolet light, methyl methanesulfonate and 4-nitroquinoline-1-oxide is not a passive process. Rather, it requires the intervention of a cellular system that processes damaged DNA in such a way that mutations result. This type of processing is referred to as "error prone repair". This process can be blocked by mutations at three chromosomal loci, lexA, recA and umuC. Mutations at the recA and lexA loci are pleiotropic. The product of these genes have been well characterized and have been shown to regulate a diverse set of functions collectively known as the "SOS functions", included among which is "error-prone repair". In contrast, mutations at the umuC locus appeared to be specific for the "error-prone" function of the "SOS response". Thus, the umuC gene product(s) is the best candidate for the protein(s) directly involved in the mutagenic process.

This thesis describes the isolation and characterization of both the umuC locus and a related locus, muc, located on the pKM101 mutagenesis-enhancing plasmid. The umuC locus was isolated by first using a combination of genetics and molecular cloning to generate a probe specific to umuC DNA. This probe was then used to identify a Charon 28  $\lambda$  clone carrying the umuC locus. Through a combination of subcloning and Tn1000 mutagenesis, a 2.2 kb region of DNA was identified which contained the information necessary to complement umuC mutations. This region of DNA was shown to code for two polypeptides with molecular weights of 16,000 and 45,000 daltons. The genes coding for these proteins were shown to be organized in an operon that is repressed by the lexA protein. Complementation of previously isolated umuC mutations revealed two complementation groups, umuC and umuD. umuC codes for the 45,000 dalton protein, and umuD, codes for the 16,000 dalton protein, and therefore both proteins are essential for "error-prone repair" in *E. coli*. Overproduction of the umuC umuD gene products on a high copy number plasmid was shown to interfere with the normal functions of the "error-prone" pathway and to be detrimental to cell growth in the presence of the recA gene.

The mutagenesis-enhancing plasmid pKM101 plays a major role in the success of the Ames test for identifying environmental carcinogens as mutagens. This plasmid can suppress mutations at the umuC locus and is hypothesized to carry a functional analog of the umuC locus (114). Recently, the region responsible for this suppression, the mucA/mucB locus, was shown to code for two polypeptides similar in molecular weight to the umuD and umuC proteins (122). This suppression is known to be recA lexA-dependent (108). To investigate the regulation of this locus, a mucB-lacZ protein fusion was constructed by *in vitro* techniques which placed the  $\beta$ -galactosidase protein under the regulation of the promoter-regulatory region of the mucA/B locus. In strains harboring this fusion,  $\beta$ -galactosidase activity was found to be induced by agents which damage DNA. Through genetic and maxicells analysis this locus was shown to be organized in an operon that is repressed by the lexA protein.

DNA sequencing analysis of the umuD gene and the beginning of the umuC gene revealed two lexA binding sites separated by only four nucleotides in the region directly preceding the translational start of the umuD gene. These lexA protein binding sequences are arranged such that when filled by bound lexA protein, they would sterically block binding of RNA polymerase to the likely promoter sequence. An open reading frame of 360 nucleotides, coding for 120 amino acids, was discovered for the umuD gene. This open reading frame terminates at tandem termination codons followed by 60 nucleotides, the translational start of the umuC gene and an open reading frame until the end of the sequenced region.

The nucleotide sequence of the beginning of the mucB gene was also determined. Comparisons of the mucB nucleotide sequence and deduced amino acid sequence with the corresponding region of the umuC sequence revealed extensive homology. 50% of the amino acid sequence was conserved as was 53% of the nucleotide sequence. Thus, the umuD/C locus and the mucA/B locus diverged from a common evolutionary precursor.

## Autobiographical Note

Name: Stephen J. Elledge  
Social Security Number: Redacted  
Date of Birth: August 7, 1956  
Place of Birth: Paris, Illinois  
Home Address and Telephone: 28 Inman Street, Apt. 6  
Cambridge, MA 02139  
(617)864-4409  
Office Address and Telephone: 56-623, Dept. of Biology  
Massachusetts Institute of Technology  
Cambridge, MA 02139  
(617)253-4721

## Education:

Secondary School: Paris High School  
Paris, Illinois

Universities: University of Illinois (1974-1976)  
University of Southampton  
Southampton, England (1976-1977)  
University of Illinois (1977-1978)  
B.Sc., Chemistry, Summa Cum Laude

Honors: James Scholar  
Illinois State Scholar and Scholarship  
Monsanto Scholarship  
Chemical Industries Council Scholarship  
Kiwanis Scholarship  
Eliot Ritchie Alexander Award  
Rotary Special Scholarship  
Phi Lambda Upsilon Award  
Eta Sigma Phi Honorary Fraternity  
Bronze Tablet University of Illinois

## Publications:

Elledge, S.J. and Walker, G.C. Proteins required for UV and chemical mutagenesis: identification of the products of the umuC locus of Escherichia coli. submitted to J. Molec. Biol.

Krueger, J.H., Elledge, S.J. and Walker, G.C. Isolation and Characterization of Tn5 Insertion Mutations in the lexA Gene of Escherichia coli. submitted to J. Bacteriol.

Elledge, S.J. and Walker, G.C. Inducibility of the muc functions of pKM101. in preparation.

## Table of Contents

	Page
Title Page	1
Acknowledgements	3
Abstract	4
Autobiographical Note	6
Table of Contents	7
List of Tables	9
List of Figures	10
List of Abbreviations	12
CHAPTER I. Introduction	13
I. Historical Overview of the SOS System	13
II. Identification of the Genes Involved in Chemical Mutagenesis	16
III. Chemical Mutagenesis	22
IV. Possible Mechanisms for "Error Prone Repair"	29
V. Plasmid Encoded Repair Functions	36
References	40
CHAPTER II. The Isolation and Characterization of the <u>umuC</u> Locus of <i>E. coli</i> and Identification of the Gene Products.	44
Introduction	44
Results	46
I. Isolation of a <u>umuC::Tn5</u> Mutation.	46
II. Isolation of a <u>umuC</u> -Specific Probe.	47
III. Isolation of a $\lambda$ Clone Containing the <u>umuC</u> Locus and Complementation by $\lambda$ - <u>umuC</u> <sup>+</sup> .	48
IV. Subcloning the <u>umuC</u> Region.	52
V. Identification of the Proteins Required for UV Mutagenesis.	58
VI. Direction of Transcription and Regulation of the <u>umuC</u> Locus.	64
VII. Identification of Two Complementation Groups and Reassignment of Previously Isolated <u>umuC</u> Alleles.	68
VIII. Complementation Properties of the <u>umuD/C</u> Locus on a High Copy Number Plasmid.	73
Discussion	81
References	87

Table of Contents (continued)	Page
CHAPTER III. Studies of the Regulation of the <u>muc</u> Locus of pKM101.	89
Introduction	89
Results	91
I. Construction of a <u>mucB-lacZ</u> Protein Fusion.	91
II. Damage-inducibility of the <u>mucB</u> gene.	95
III. Genetic Analysis of the Regulation of <u>mucB</u> .	95
IV. Effect of <u>recA430</u> on <u>mucB</u> Expression.	102
V. Analysis of the Regulation of <u>mucA</u> and <u>mucB</u> Using Maxicells.	106
Discussion	110
References	114
CHAPTER IV. The Nucleotide Sequence of the <u>umuD</u> Gene and Comparisons Between Partial Nucleotide and Amino Acid Sequences of <u>umuC</u> and <u>mucB</u> .	115
Introduction	115
Results	117
I. The Regulatory Region of the <u>umuD/C</u> Operon.	117
II. Open Reading Frames	125
III. Comparison of the <u>umuC</u> and <u>mucB</u> Nucleotide and Deduced Amino Acid Sequences.	128
Discussion	131
References	134
CHAPTER V. Materials and Methods	135
I. Bacterial Strains, Media and Chemicals	135
II. Plasmid Constructions	135
III. DNA Preparation	142
IV. Electrophoresis of DNA and Southern Blotting	143
V. Preparation of <sup>32</sup> P-labeled DNA	143
VI. Hybridization	144
VII. UV Mutagenesis and Survival	144
VIII. Maxicells and Electrophoresis of Proteins	145
IX. Tn1000 Mutagenesis	146
X. $\beta$ -Galactosidase Assays	146
XI. DNA Sequencing	147
References	148



## List of Tables

		Page
Table 3-1.	Induction of $\beta$ -galactosidase in strains carrying pSE200 by various agents.	96
Table 5-1.	Bacterial Strains	136
Table 5-2.	Plasmids	138

List of Figures		Page
Figure 2-1.	Identification of a <u>umuC</u> probe and analysis of insertions in and near the <u>umuC</u> locus.	49
Figure 2-2.	Identification of $\lambda$ <u>umuC</u> <sup>+</sup> clones in transient merodiploids.	53
Figure 2-3.	The restriction map of the <u>umuC</u> region of <u>E. coli</u> and the location of <u>umuC::Tn1000</u> insertions.	56
Figure 2-4.	UV mutagenesis and survival in a <u>umuC</u> <sup>-</sup> strain carrying a <u>umuC</u> <sup>+</sup> plasmid.	59
Figure 2-5.	Identification of the proteins encoded by the <u>umuC</u> locus of <u>E. coli</u> .	62
Figure 2-6.	The position and transcriptional organization of the <u>umuC</u> and <u>umuD</u> product.	65
Figure 2-7.	Repression of the synthesis of <u>umuC</u> and <u>umuD</u> protein synthesis by the <u>lexA</u> protein in maxicells.	69
Figure 2-8.	Complementation of UV mutagenesis of <u>umuC</u> and <u>umuD</u> mutants.	71
Figure 2-9.	Schematic illustration of the construction of pSE117.	75
Figure 2-10.	Complementation of UV mutagenesis and survival of a <u>umuC36</u> mutant by pSE117, a high copy number plasmid carrying the <u>umuC/D</u> locus.	77
Figure 3-1.	Schematic illustration of the construction of pSE200.	92
Figure 3-2.	Kinetics of UV induction of $\beta$ -galactosidase activity in various genetic backgrounds containing pSE200.	97
Figure 3-3.	Analysis of the kinetics of induction of $\beta$ -galactosidase activity in strains bearing mutations at the <u>recA</u> locus.	100

Figures (continued)	Page
Figure 3-4. Analysis of the differential effects of the <u>recA430</u> allele on the induction of the <u>dinB::Mu d(Ap<sup>R</sup>,lac)</u> fusion located on the chromosome (GW1039) or on a multicopy plasmid (pGW510).	103
Figure 3-5. Repression of the synthesis of the <u>mucB'-lacZ'</u> protein fusion and the <u>mucA</u> product by the <u>lexA</u> protein in maxicells.	107
Figure 4-1. Restriction Map of the <u>umuC/D</u> locus showing the positions of peptides and the direction and extent of sequencing.	118
Figure 4-2. A DNA sequencing gel of a fragment containing the operator-promoter and translational start of the <u>umuD</u> gene.	120
Figure 4-3. A. Comparison of the <u>lexA</u> binding site of the <u>umuC/D</u> operon with sequences of other <u>lexA</u> binding sites. B. Functional organization of the <u>umuD/C</u> regulatory region. C. Comparison of ribosomal binding sites of the <u>umuD</u> and <u>umuC</u> genes with the 16S rRNA sequence.	123
Figure 4-4. Nucleotide sequence of the <u>umuD</u> gene and the beginning of the <u>umuC</u> gene along with the deduced amino acid sequence.	126
Figure 4-5. Comparison of the nucleotide sequence and the deduced amino acid sequence of the amino terminal end of the <u>umuC</u> and <u>mucB</u> genes.	129

## Abbreviations

DNA =	deoxyribonucleic acid
UV =	ultraviolet light
J =	joule
SDS =	sodium dodecyl sulfate
kb =	kilobase
EDTA =	ethylenediaminetetraacetic acid
kd =	kilodalton
Km =	kanamycin
Sp =	spectinomycin
Ap =	ampicillin
Tc =	tetracycline
Cm =	chloramphenicol
MW =	molecular weight
Def =	defective
Ind =	inducible
EMS =	ethyl methanesulfonate
MNU =	N-methyl-N-nitrosourea
MNNG =	N-methyl-N'-nitro-N-nitrosoguanidine
ENU =	N-ethyl-N-nitrosourea
MMS =	methyl methanesulfonate
BPDE =	benzpyrene diol epoxide
NCS =	neocarzinostatin
4NQO =	4-nitroquinoline-1-oxide
AP =	apurinic or apyrimidinic
CPPE =	epoxycyclopentapyrene
AAF =	N-2-acetyl aminofluorine

## Chapter I

### Introduction

#### I. A Historical Overview of the SOS System

The concept of an inducible error-prone repair system in E. coli arose from the early experiments of Weigle (1) who observed that both the survival and mutagenesis of UV-irradiated  $\lambda$  phage was increased by pre-irradiation of the bacterial host. This phenomenon was termed Weigle-reactivation and Weigle-mutagenesis and was found to require the recA<sup>+</sup> and lexA<sup>+</sup> (2) genotype in the host. UV mutagenesis of the bacterial chromosome was also found to be recA<sup>+</sup> lexA<sup>+</sup>-dependent (3) suggesting that the same error-prone DNA repair system responsible for Weigle-reactivation was also responsible for UV mutagenesis in E. coli. The suggested inducibility of this system was supported by a number of observations. First, the induction of Weigle-reactivation is inhibited by the presence of chloramphenicol during the preinfection incubation (4,5) indicating a need for de novo protein synthesis. Secondly, a comparable degree of reactivation is observed without UV by infecting a tif1 host cell that preincubated at 42°C (2). This induction is blocked by a lexA3 (Ind<sup>-</sup>) mutation. Finally, low UV doses which are normally submutagenic produce mutagenesis when applied to a tif1 cell preincubated at 42°C (6).

In addition to error-prone repair, treatment of E. coli with agents which damage DNA or block DNA replication cause induction of a number of physiologically diverse phenomena (For a review see 7). These include the induction of the recA protein (protein X) (8), induction of certain lambdoid prophages (7,14), inhibition of respira-

tion (9,10), inhibition of DNA degradation (11), stable DNA replication (12,13), and inhibition of cell septation leading to filamentation (15,16). The idea that the recA and lexA gene products coordinately regulate these seemingly unrelated processes was put forth as the "SOS" hypothesis by Radman (17). These processes are now collectively called the SOS functions because they appear to increase the probability that the cell will survive (7,17).

A great deal of knowledge has been discovered in the last 10 years concerning the molecular mechanics of the regulation of the SOS functions. In 1975, Jeff Roberts and his collaborators showed that the lambda repressor was cleaved in vivo following UV irradiation (18). They later went on to demonstrate in a very controversial paper that the protease responsible for this cleavage was the recA protein (19). This explained the induction of prophage  $\lambda$  by DNA damage. Little et al. (20) then showed that the recA protein cleaved the lexA protein, which was subsequently shown by Little et al. and Brent and Ptashne to repress both itself (22,23) and the recA gene (20,21). These experiments explained the induction of the recA protein in response to DNA damage. DNA damage produces a signal (single-stranded DNA ?) that activates the protease function of the recA protein which proceeds to cleave the lexA repressor. As levels of lexA protein drop, the recA gene is derepressed and thus the recA gene product is induced. As soon as the activating signal disappears, presumably due to the repair capacities of the SOS functions, the recA protein is no longer active proteolytically and the lexA protein levels rise until the lexA protein begins to repress its own synthesis - a finely-tuned

regulatory circuit. The lexA gene product was then postulated to control other genes involved in the SOS response. Furthermore, the lexA3 allele proved to produce an uncleavable repressor protein explaining its dominance and its ability to block induction of all the SOS functions (24).

Searching for a means to investigate the regulation of other genes induced in the SOS response, Kenyon and Walker devised a clever experiment to approach this problem with great generality. They isolated a random series of Mu d(Ap<sup>R</sup>, lac) insertions in the E. coli chromosome which fuses the lac operon to the promoter of any gene in which the Mu d(Ap<sup>R</sup>, lac) inserts. Thus in one step an insertion mutation is created as well as an easily assayable operon fusion which is under the regulation of the promoter-regulatory region of the mutant gene or transcriptional unit bearing the insertion (2,6). They found a number of insertions in which  $\beta$ -galactosidase activity was induced in response to DNA damage. The loci containing these mutations were termed din (damage-inducible) genes (25). Other damage-inducible genes were subsequently identified by Kenyon et al. and others and now the number of inducible loci stands at 13 (27). Among the functions thus far identified for these genes are uvrA (28), uvrB (29,30), umuC (31), sfiA (32), and himA (33). Damage-inducible plasmid-localized genes identified thus far are the colicin E1 gene on colE1 (34) and the muc gene on pKM101 (this thesis).

The existence of lac operon fusions greatly facilitated genetic analysis of the transcriptional control. Introduction of a lexA(Def),

formerly spr, mutation into the various din strains rendered expression constitutive at the induced level even in the presence of a recA<sup>-</sup> mutation. This suggested that lexA directly repressed the transcription of the din genes. In vitro studies of inhibition of transcription and DNA footprinting utilizing several cloned promoters of damage-inducible genes and purified lexA protein have shown that the lexA protein does indeed repress transcription of these genes by directly binding to the promoter region (27,35). Furthermore sequences of the promoter regions have defined a canonical lexA binding sequence (27).

## II. Identification of Genes Involved in Chemical Mutagenesis

As mentioned earlier, chromosomal mutagenesis, Weigle-reactivation and Weigle-mutagenesis were known to depend upon the recA<sup>+</sup> lexA<sup>+</sup> genotype. These mutations were identified while screening for phenotypes unrelated to their involvement in mutagenesis; recA was discovered in 1965 by Clark and Margulies while searching for genetic loci involved in recombination (36); lexA (Ind<sup>-</sup>) mutations were first isolated in 1966 by Howard-Flanders and Boyce for their increased sensitivity to UV light and ionizing radiation (37). Only upon subsequent characterization were they discovered to be deficient in mutagenesis. In 1977, Kato and Shinoura undertook the task of isolating mutations in genes involved in mutagenesis. By screening 30,000 EMS mutagenized clones, they isolated 6 umu (unmutable) mutations characterized by their inability to revert the his-4 ochre mutation to His<sup>+</sup> in the presence of UV irradiation. The six mutations fell into three



categories on the basis of UV sensitivity and recombinational proficiency. The first class, umuA, mapped to the lexA locus, umuB mapped to the recA locus and was recombinational deficient, while a third and novel class of mutations, umuC were found to map at 25.5 minutes (38).

The three umuC mutations isolated by Kato had very similar phenotypes: i) they were nonmutable by UV light and MMS, ii) they were partially deficient in Weigle-reactivation (totally deficient in a uvrA<sup>-</sup> background), iii) they were slightly UV sensitive, and iv) they were recombinationally proficient. Unlike recA and lexA mutations which are deficient in all SOS functions due to their uninducibility, umuC mutations affected only the error-prone repair functions while allowing other SOS functions to proceed normally. Although this does not rule out a regulatory role for umuC, there is no evidence to suggest a regulatory role and thus far the umuC protein(s) appears to be the best candidate for the gene product(s) (5) involved mechanistically in the error-prone repair process. A year later Steinborn isolated uvm mutations with the same phenotype and map position as umuC and these are most surely alleles of umuC (39). In 1980, a Mu d(Ap,lac) insertion in the umuC locus was isolated by screening 17,000 independent insertions in E. coli for loss of UV mutability (31). This umuC::Mu d(Ap,lac) insertions, umuC121 (40), had the same phenotype as the umuC mutants isolated by Kato, suggesting that perhaps this was the null phenotype. Of course this did not prove that point since insertion mutants, like nonsense mutants, may easily leave partially functional truncated polypeptides.  $\beta$ -galactosidase

expression was shown to be inducible 15 fold by DNA damaging agents including UV in the umuC:Mu d(Ap,lac) fusion strain. Genetic analysis of this induction revealed a recA<sup>+</sup>lexA<sup>+</sup>-dependence. Furthermore,  $\beta$ -galactosidase was produced constitutively in the presence of a lexA(Def) allele regardless of the recA allele suggesting that the lexA protein was the direct repressor of the umuC gene. Thus it appeared as if the puzzle of the genetic dependence of UV mutagenesis was solved; cells bearing mutations at the recA or lexA locus were non-mutable because the umuC gene product(s) were not induced. This explanation predicts that cells lacking the lexA gene product, i.e. lexA(Def), would no longer require a functional recA gene product for mutagenesis because they would already be producing the umuC protein(s) constitutively. However, when recA lexA(Def) cells were tested for mutagenesis, they were found to be UV nonmutable (J. Kreuger per. comm., 27). This suggests a second role for the recA protein in UV mutagenesis. Among the possible roles are, i) a direct mechanistic involvement in the mutagenic process, ii) a positive effector of the expression a gene directly involved in mutagenesis but not repressed by the lexA protein, and iii) a positive activator of an activity involved in mutagenesis, perhaps even the umuC protein(s).

Other experiments involving the tif1 mutation have also implied a second role for the recA protein in mutagenesis. Tif1 is an allele of recA which shows induction of SOS functions at 42°C including a substantial increase in spontaneous mutagenesis (7,41). The tif1 protein has been examined in vitro by Phizicky and Roberts who found this protein to require a lower concentration of triphosphates and a

substantially lower concentration of polynucleotides than the recA<sup>+</sup> protein to activate its protease activity (42). Presumably at 42°C the concentration of these cofactors is increased to a level high enough to activate the protease activity of the tif1 protein. If the only function of the recA protein protease activity is to cleave the lexA protein, then in the absence of the lexA gene product, lexA(Def), recA<sup>+</sup> and tif1 should be equivalent at all temperatures. By examining the effects of tif1 and recA<sup>+</sup> on spontaneous mutagenesis in a lexA(Def) background, Blanco et al. (43) found that the spontaneous mutation rate was much higher in a tif1 lexA(Def) at 42°C than at 30°C. Furthermore, they found the spontaneous mutation rate of a tif1 lexA(Def) strain was much higher than a recA<sup>+</sup> lexA(Def) strain at 42°C and that the recA<sup>+</sup> lexA(Def) strain showed no temperature effect regarding mutagenesis. Thus the increase in spontaneous mutagenesis rate appears to correlate with the proteolytic activity of the tif1 protein or at least with the conformation of the recA protein when it is active proteolytically. Blanco et al. interpreted their results to mean that not only is the recA protein needed for mutagenesis but that it must also be proteolytically active. This is the simplest interpretation of the data but it must also be added that this interpretation relies heavily on the assumption that the lexA(Def) allele they used, spr51, is truly the null allele of lexA.

An experiment performed by Sekiguchi et al. also sheds light upon the role of the recA protein in mutagenesis. They developed a technique which allowed them to introduce proteins into cells by plasmolysis while retaining a large degree of cell viability (44).

This allowed them to introduce proteins from  $UmuC^+$  strains into umuC backgrounds and assay for UV mutability. They found that extracts prepared from DM1187 (tif1 lexA(Def)) would not complement a strain containing the umuC36 mutation unless DM1187 was grown at 42°C or induced with UV light before making the extract (45). This result is surprising because the host cells, containing the umuC36 mutation, had been induced with UV light and therefore contained proteolytically active recA protein while the source of the extracts were lexA(Def) and should therefore be producing the umuC protein(s) constitutively. There should be no lack of the umuC protein(s) in the extracts and no lack of activated recA protein in the host cells. Why then the lack of complementation or the absence of inducing treatment? Perhaps the recA protein when proteolytically active stabilizes the umuC product from degradation in the extracts or active recA protein may solubilize the umuC protein(s) so that it remains in the supernatant during preparation of the extracts. The answer to this question must await further analysis of the biochemical properties of the umuC protein (5). However, these experiments do point to a more complicated role for the recA protein in the mutogenesis process than was previously suspected.

Other mutations have been isolated that affect SOS repair. For example, mutations in the uvrA/B/C, recL (uvrD), and recF genes have been shown to alter the cells' ability to perform Weigle-reactivation. The uvrA/B genes, which are responsible for excision repair of bulky damage to DNA, have recently been found to be inducible by UV irradiation thus explaining their requirement in Weigle-reactivation

(28,29,30). The recL (uvrD) gene has recently been identified as the helicase II protein (46). It has been implicated in a number of repair pathways including excision repair. Apparently thymine dimers, which are one of the major photoproducts of UV irradiation, are excised at a much slower rate in a uvrD<sup>-</sup> background (47). Since uvrA/B dependent incision at pyrimidine dimers has been shown to nick on both sides of a pyrimidine dimer (48) leaving a 12 base dimer-containing fragment annealed to the undamaged strand. Perhaps uvrD is needed to unwind this 12-mer so that it will be removed and repair synthesis may begin. Recent unpublished data have indicated that the uvrD protein is also damage inducible 2 to 3 fold (P. Pasz, personal comm., E. Siegle, personal communication).

The recF mutation differs from uvrA/B/C and uvrD mutations in that it totally blocks Weigle-reactivation in a uvrA background (38) but does not block cellular mutagenesis nor Weigle-mutagenesis (49). The problem of determining the role of the recF gene product in Weigle-reactivation is particularly difficult because recF mutants show reduced levels of expression of a number of SOS functions including recA protein induction and possibly induction of prophage  $\lambda$  (50,51). It is possible that recF mutations affect Weigle-reactivation by reducing the extent of induction of the uvrA, uvrB and umuC gene products. Clark and collaborators have shown that the defect produced by recF mutations is not due merely to a lowering of recA protein levels since recA<sup>OC</sup> (operator constitutive) mutations do not suppress this RecF phenotype (52). However if the effect of recF is actually to lower the level of inducing signal or a delay in

its accumulation, then a recA<sup>OC</sup> mutation would not be expected to alter the RecF phenotypes. An interesting experiment would be to look at the level of Weigle-reactivation in a tif1 recF double mutant by shifting from 30°C to 42°C. If recF blocks induction by delaying or decreasing the inducing signal, then tif1 should be able to overcome this defect and restore Weigle-reactivation while remaining proficient in thermal induction of prophage  $\lambda$ . If this were true, then that would imply a mechanistic involvement of the recF gene product in Weigle-reactivation.

### III. Chemical Mutagenesis in E. Coli

Chemical mutagenesis in E. coli is thought to occur through two distinct mechanisms, mispairing (misreplication) and misrepair (60). Misreplication involves erroneous base-pairing directed by base analogs and mutagen-modified bases. Mutagens in this class are called direct mutagens (61) because they rely solely on the replication machinery normally present in the cell for their mutagenic potential and hence are recA<sup>+</sup> lexA<sup>+</sup> and umuC<sup>+</sup>-independent (63). Direct mutagens are highly base specific (62) and some examples of this class are ENU, MNU, MNNG and EMS (63,64). The misrepair mutagens are thought to create non-pairing lesions containing little or no template information. These lesions cause blockage of DNA replication and consequently the induction of the SOS error-prone repair system. Mutagens of this class (e.g. MMS,  $\gamma$ -irradiation, UV, BPDE, neocarzinostatin and 4-NQO (64,70)) are called indirect mutagens because of their dependence on the induction of the SOS repair functions (61) and are therefore recA<sup>+</sup> lexA<sup>+</sup> and umuC<sup>+</sup>-dependent.

A major issue in the field of mutagenesis has been whether or not mutations created by indirect mutagens are targeted. I will define targeted mutations as those mutations that arise at the same base as was modified or damaged by the mutagen itself. Evidence supporting the presence of untargeted mutagenesis has come from a number of experiments. Ishikawa-Ryo and Kondo observed that mutagenesis of undamaged  $\lambda$  was increased when infecting a host preinduced for SOS functions with UV (65). Furthermore, tif1 strains when grown at 42°C show a mutator phenotype (6,15), i.e. they increase the level of spontaneous mutagenesis by a factor of 4 at 42°C and a factor of 20 at 42°C in the presence of adenine (68). Thermal treatment of a tif1 strain after exposure to low, uninducing doses of UV or  $\gamma$ -irradiation, however, enhances the yield of radiation induced mutations far above the sum of the frequencies of induced mutations observed in controls treated only with heat or only with radiation (6,66,67). Thus upon induction of SOS repair, a certain amount of mutagen independent mutagenesis (but not necessarily lesion independent) occurs. However, Witkin has estimated that less than 1% of UV induced mutations would be untargeted by her calculations (she was using thermal induction of tif1 as her fully induced control: treatment at 42°C plus adenine would have increased her background 5 to 6 fold (43)). Nevertheless, it appears that mutagen targeted mutagenesis is the dominant pathway of indirect mutagenesis.

Strong support for the targeted mutagenesis hypothesis arrived with the appearance of a system designed to measure the base pair

specificity of different mutagens. In 1977 Coulondre and Miller (69) developed a forward mutation system using the lacI gene which allowed them to monitor the mutational specificity of a particular mutagen. Their assay, which utilized over 70 characterized nonsense mutations, was able to detect all transitions and transversions with the exception of the A•T to G•C transition. They found that each mutagen had a unique profile of site preferences as well as base pair specificity. For example, 4-NQO gave 92% G•C to A•T transition and favored no particular site while UV showed a preference for the G•C to A•T transition (61%) but also showed all classes of transitions and transversions, including an extrapolated 10% tandem double base pair changes. UV also showed a very characteristic pattern of mutational site preference within a given set of transitions or transversions. The fact that each mutagen yields a different pattern and specificity of mutagenesis has been interpreted to mean that each mutagen creates its own unique set of targeted premutational lesions which are fixed into mutations by the error-prone repair system of E. coli.

Foster et al. (70) have analyzed the UV-generated spectrum of Coulondre and Miller as well as their own data for NCS (71) and have shown that by analyzing the frequencies of UV induced mutations per site, regardless of their type, it becomes apparent that the mutations can be grouped into 2 classes of events: apparently random, low frequency occurrences (LFOs) that account for one third of the mutations generated and non-random high frequency occurrences (HFOs) that account for the rest. Furthermore, of the 23 HFOs, all are at sites of adjacent pyrimidines which are known to be involved in the primary photo-



products of UV irradiation. Of the remaining 41 LFOs, 26 are adjacent pyrimidines and 15 are pyrimidines flanked by purines indicating that adjacent pyrimidines are not enough to make a UV generated HFO. The LFOs fall into a Poisson distribution among available sites and could be considered to be untargeted events. LFOs from NCS, like UV, contain examples of all 5 base pair changes than can be monitored but, unlike UV, NCS mutates only 32 of the 64 possible sites. Therefore, even the LFOs are probably targeted events. Foster et al. have suggested that since among LFOs all base pair changes are seen, the lesion producing LFOs may be the same for several mutagens as opposed to HFOs which appear specific. An example of a non-informational lesion that several mutagens could share would be an apurinic or apyrimidinic site which are known to be mutagenic (72).

Although the lacI system is generally better suited to monitor transitions and transversion to A•T pairs than to G•C pairs, approximately 8 sites to 1, it still appears even after normalization of the difference in available sites that SOS dependent mutagens produce A•T base pairs preferentially. It may very well be true that each mutagen creates its own special lesions and that each of these lesions interacts with the cells' repair system differently to instruct it to place a particular incorrect base across from it thereby generating the spectrum of mutagenesis observed for that mutagen. This hypothesis seems unlikely in light of the fact that most SOS dependent mutagens create lesions which block normal polymerization completely (73,74) and at that point apparently have very little informational content left with which to instruct the polymerase. I would like to

present a model which would account for the observed spectra yet require little or no informational content on the part of the damaged base. This model would require one assumption, that the SOS repair system react to non-informational lesions by inserting an adenine across from them at a high frequency and the other bases at a low but discernible frequency independent of the exact nature of the lesion. Thus the specificity of insertion lies completely with the polymerase and the SOS system. The mutagen in question would be presumed to create several different lesions (i.e. interactions with different bases) and, in some cases, one lesion preferentially. Let us use aflatoxin B<sub>1</sub> as our example. Aflatoxin B<sub>1</sub> is known to create an adduct at the N7 position of guanine and that this lesion accounts for 90% of its lesions. Let us assume that the other 10% of the lesions are distributed randomly among A, T and C. If A is inserted 20% of the time across from a lesion and C, T and G each 3.3% of the time, then among 90 insertions resulting from the N7G lesions we would generate 81 G.C to T.As, 3 G.C to A.Ts, 3 G.C to C.G and 4 would be unchanged. At the 10% of the adducts other than G we would produce 3 A.T to T.A from insertion across from A, 3 G.C to A.T from insertion across from C and perhaps one or fewer A.T to C.G and A.T to G.C for insertion other than A across from A or T adducts. This would yield the final approximate percentages for the mutations of 86% G.C to T.A, 6% G.C to A.T, 3% A.T to T.A, 3% G.C to C.G, and one or less percent of A.T to C.G and A.T to G.C. The actual numbers are 89% : 6% : 3% : 0 : 0.5% : ?. Of course, these numbers were chosen to mesh well with the example, but they do serve to illustrate the point that spectra can be generated by considering the sites and frequency of mutagen

interacting with the DNA as the sole variable. In this model the HFOs would be considered to be mutations where adenine was inserted across from the major lesions, G in this example. LFOs would be composed of all non A inserts plus all A inserts across from minor lesions. This would account for the targeted nature of LFOs as well as their ability to show all base pair changes. Predictions from this model are:

1) that all SOS dependent mutagens show a strong preference for transitions and transversions to A•T base pairs (e.g. MMS, tif1 mutator effect) and 2) that mutagens which produce a particular base pair change predominantly, like G•C to A•T for 4-NQO, will be shown to interact with the base across from A initially, i.e. C in this case.

Recent experiments have generated data which support this model.

Transfection of apurinated  $\phi$ X174 amber mutants into SOS induced hosts (apurinic sites are SOS dependent mutagenic lesions (75) and sequence analysis of the revertants have shown a strong preference for insertion of A across from the apurinic lesion (Larry Loeb, unpublished result). The lacI spectrum of tif1 has recently been shown to be dominated by G•C to TA transversions (Jeffrey Miller, personal communication). The lacI spectrum for MMS is currently under way (Jeffrey Miller, personal communication). Unfortunately, 4-NQO is thought to interact primarily with G instead of C (76). However, this model is based on non-coding lesions, if a particular lesion retains some ability to code or instruct the polymerase, then A need not be the major insert. Models hypothesizing how different damaged bases could still instruct a polymerase have been presented by Topol and Fresco (77).

The overriding problem in deciphering the mutagenic specificity of a particular mutagen is that each mutagen creates a number of different lesions each differing in frequency of occurrence, site specificity and mutagenic potential. Thus each lacI spectrum is a weighted superposition of a number of separate spectra, each specific for a particular type of lesion filtered through the various repair systems present in E. coli. In order to understand the spectrum of a given mutagen we need to first identify the lesions produced by the mutagen and then discover the mutagenic specificity of each lesion. This would entail preparing molecules that contained only one defined lesion and then measuring its specificity of mutagenesis. Although this has yet to be completed for any mutagen, great progress has been made in this area. Loeb and his colleagues have shown that apurinic and apyrimidinic lesions in single stranded phage  $\phi$ X174 are mutagenic (78) and that the formation of these mutations depends upon the SOS repair system (75). This is an important finding since AP sites are a possible common intermediate in a number of different lesions.

Identification of the chemical nature of lesions produced by known carcinogens is an area of intense research. Correlations of mutational specificity and sites of damage have been made for AAF, aflatoxin B<sub>1</sub>, BPDE and CPPE which interact with Gs and primarily cause G•C to T•A transversions (79,80,81). There is also a strong correlation in the lacI system between the mutation rate in vivo and UV induced base damage in vitro. Brash and Hazeltine have measured this frequencies of pyrimidine-pyrimidine cyclobutane dimers and 6-4 photo-products at specific sites in the lacI gene. They have shown a strong

correlation between amounts of UV induced damage (pyrimidine-pyrimidine cyclobutane dimers and 6-4 photoproducts) with mutability (82). Of course, the lacI spectrum they used for comparison was performed in a uvrA<sup>+</sup> background (69), a fact they fail to mention. The most mutable sites are TC and CC which also have a high proportion of 6-4 photoproduct. Unfortunately it is impossible to measure the true spectrum since the TA to CG transition is not measured in the lacI system. An analysis of specificity in a missense system is needed to determine the TA transition as well as the true mutability of a TT dimer.

#### IV. Possible Mechanisms for "Error-Prone Repair"

The culminating question in the field of error-prone repair is "What is the mechanism of action of SOS repair; where and how does it work?" This is a difficult question to approach in an organized manner because there are many theories and little hard evidence. In order to understand how mutations might arise, it is important to first understand how they are prevented. The fidelity of the E. coli replication machinery is such that it is estimated that the average base pair undergoes a mutational alteration only once in every  $10^8 - 10^{10}$  replications of the genome (83,84). This tremendous fidelity is due to three groups of molecular processes which are involved in the maintenance and accurate replication of genetic information. The first group of processes are involved in the maintenance of the genetic information. These are error-free pathways of DNA repair and include excision repair, photoreactivation of cyclobutane dimers, adap-

tive repair (dealkylation of  $O^6$  alkyl guanine), uracil excision repair (uracil-DNA glycosylase, AP endonucleases ) and post-replicative recombinational repair. With the exception of recombinational repair, all of these processes are specific for a certain type of common lesion. None of these processes are needed for SOS repair and will not be discussed further in this introduction.

The second group of processes involve base selection at the replication fork. The high fidelity of prokaryotic polymerases arises from a low frequency of base misinsertion (base selection error) combined with an editing or proofreading activity that excises mismatches. The error rate of base selection have been estimated to be between  $10^{-4}$  and  $10^{-5}$  (86) and proofreading increases that fidelity by a factor of 10 to 100 (85,86). The error rate of misincorporation can be altered by increasing the concentrations of an incorrect dNTP over the corrected dNTP, i.e. doubling the concentration of dATP relative to the other dNTPs will double its misincorporation frequency simply by mass action (87). Kunkel et al. have shown that accessory proteins such as SSB may also increase the fidelity of polymerization (88).

The effects of base selection, proofreading and single-stranded binding protein on the fidelity of replication work to reduce the error frequency to approximately  $10^{-6}$  to  $10^{-7}$ . The remaining increase in fidelity results from the action of the third set of processes, postreplicative methyl-directed mismatch repair (89). Mismatch repair is dependent upon the dam, mutH, mutL, mutS and uvrD genes. Briefly, the system is thought to work in the following manner: The product of

the dam gene is a methylase which methylates the N<sup>6</sup> position on the adenine ring at GATC sequences in DNA. During replication, the parental strand is fully methylated while methylation of the daughter strand lags behind the replication fork. Thus newly synthesized DNA is transiently hemimethylated and hemimethylated DNA is the substrate for mismatch repair. Presumably the products of the mutH, L and S genes and possibly uvrD (helicase II) locate base mismatches and initiate correction of the mismatch using the information on the parental (methylated) strand as the correct information. This mismatch repair system contributes another two to three orders of magnitude to the fidelity of replication (90).

One might anticipate that SOS would exert its influence by introducing a novel infidelity component or by interferring with an existing component of the fidelity mechanism. For example, induction of SOS could alter the dNTP pools to produce a mutagenic imbalance of dNTPs. This is not the case, as dNTP pools are unchanged after induction of SOS except for a two fold increase in the concentration of dATP (91). Another possibility would be that induction of SOS inhibits the mismatch repair system. Radman claims to have tested this hypothesis using  $\lambda$  heteroduplex mismatches infected into SOS induced cells and found no change in mismatch repair (M. Radman, personal communication). Unfortunately, the data for this is unpublished and it is difficult to accept without first analyzing exactly how the experiment was performed.

The most popular model for SOS repair is the dimer bypass hypothesis (92,93). Evidence has suggested that polymerases stop when they encounter a dimer both in vivo (93) and in vitro (94). The hypothesis then is that SOS induced cells contain a DNA polymerase activity that polymerizes beyond dimers and creates mutations at these sites due to the inability of the dimer to base pair. There exists several pieces of evidence that support this hypothesis. First, Weigle-reactivation and Weigle-mutagenesis have been observed for several single-stranded DNA phages (55,56,57,58) and this has been shown to be multiplicity independent (57,58). This Weigle-reactivation is recA<sup>+</sup> lexA<sup>+</sup> (58) and umuC<sup>+</sup>-dependent (A.J. Clark, personal communication). Being single-stranded, these phage ( $\phi$ X174, M13, F1) are impervious to excision or recombinational repair and thus the only way to circumvent the lesion is dimer bypass. Second, a sizable fraction of UV induced mutations are tandem double base changes at potential dimer sites (69). The final piece of supporting evidence is the study of Caillet-Fauquet et al. which showed that  $\phi$ X174 A am 18 phage that were UV irradiated were able to replicate more of their DNA if they infected SOS induced cells rather than uninduced cells (93).

There are several criticisms of the evidence supporting this model. First, Weigle-reactivation of single-stranded phage is a relatively weak effect, 5 to 10 fold. Of course, the stability of a single-stranded molecule awaiting bypass may limit the degree of reactivation observable. The other major criticisms concern the biochemical evidence of bypass by Caillet-Fauquet et al. (93). The major criticisms are: 1) That they failed to show that the increased repli-



cation of their phage was a continuation of the initial replication event, i.e. secondary initiation of replication at random sites could increase the sedimentation properties of the molecules; 2) That they failed to show that molecules in the completely replicated fraction of the gradient actually contained dimers; and 3) That they did the entire experiment at an m.o.i. of 3 which means that recombination could have played a role in their "bypass". These criticisms are crucial to any proof of bypass. Perhaps the major importance of the Caillet-Fauquet et al. paper was not the evidence for dimer bypass it provided but that it was the first attempt to measure anything to do with error-prone repair biochemically.

Even the existence of an ability to bypass dimers would not necessarily imply the existence of a DNA polymerase activity capable of bypassing dimers. For example, Das Gupas and Poddar have suggested that recombination with the host chromosome could explain Weigle-reactivation of single-stranded phage (57). Another formal possibility is the "graftase" model mentioned by Elledge and Walker whereby a small oligonucleotide is grafted across the lesion, dimer, and allows replication to proceed (95). Neither of these models take into account the specificity of different mutagens, other than leaving it up to the recombinational machinery.

The dimer bypass hypothesis led directly to the search for the error-prone polymerase. However, before pursuing this topic further, it should be mentioned that the term "error-prone repair" may very well have outlived its historical relevance. In other words, it is

probably a complete misnomer. There is no real evidence that un-targeted mutagenesis exists. Certainly all chemical mutagenesis thus far examined is targeted. The low level mutator activity associated with tif1 induction of SOS may very well result from pre-existing, "cryptic", non-informational lesions which are normally repaired by some error-free pathway such as excision repair. Such a lesion might be an AP site which has error free pathways for repair in double-stranded DNA (53,54) but which is mutagenic when processed by SOS repair (73). Furthermore, the spectrum of tif1 mediated mutagenesis appears to be non-random in the lacI system (J. Miller, personal communication) suggesting the existence of cryptic lesions in the absence of DNA damaging agents. Therefore, rather than thinking of mutations occurring as a result of error-prone repair, we should instead envision them as being generated by replication across from altered bases (76).

The search for the error-prone polymerase was fueled by the rumor that Radman's group had found an error-prone polymerase activity in SOS induced E. coli crude extracts (96). However, they were unable to purify it (91). An error-prone polymerase was finally partially purified from induced E. coli by Lackey *et al.* (97). Unfortunately, this activity, which appeared recA<sup>+</sup> and lexA<sup>+</sup>-dependent (probably because of its purification properties), was a derivative of Pol I and Pol I mutants have been shown to be completely proficient in SOS repair (7).

Evidence for the direct involvement of DNA polymerase III, dnaE, in the process of SOS repair was presented by Bridges and Mottershead. Using a dnaE ts mutant in a uvrA background, they showed that DNA polymerase III activity was needed for fixation of mutations because at 42°C all mutations remained photoreversible but lost photoreversibility at 34°C (98). They suggest that a cofactor is induced by SOS treatment (umuC/D ?) which confers an "error-prone" activity upon DNA polymerase III and thus allows polymerization past dimers. When a polymerase encounters a dimer, it stalls and begins turning over dNTP to dNMP and PPi (99). This turnover has been interpreted to be the polymerase inserting a nucleotide and then excising the nucleotide with its editing function. A plausible activity for the hypothetical cofactor would be to inhibit the proofreading function of DNA polymerase III.

If SOS repair is indeed mediated by an alteration of DNA polymerase III's properties, then one might expect to find mutations located at the dnaE locus, or other loci that code for proteins present in the holoenzyme, that alter error-prone repair. This class of mutations would be expected to be UV hypermutable and independent of SOS induction. Another class of mutations one would expect would be UV nonmutables. Bridges and Mottershead found a dnaE (polC) ts revertant which had reduced UV mutability (100). Unfortunately their original mutant was in an E. coli B strain and did not transfer properly into E. coli K 12 (P. Pang, unpublished result). It may well have been a double mutant. Many revertants of dnaE ts's are second site revertants in dnaQ (mutD). (R. Maurer, unpublished results).

MutD is a strong spontaneous mutator which produces a substantial increase in all classes of base pair changes and frameshift mutations (84,102). Recently it has been conclusively shown that mutD5 is an allele of dnaQ and that the product of the dnaQ gene is  $\epsilon$ , a component of the DNA polymerase III holoenzyme (M. Sekiguchi, personal communication). Furthermore, dnaQ ts mutations become mutators at intermediate temperatures (103). One attractive model suggested by Topal and Fresco (104) is that the dnaQ protein is involved in proofreading and that mutations in dnaQ somehow inhibit the 3'-5' exonuclease activity of DNA polymerase III from excising mispaired bases. This may not be so, however, because mutD can increase mutation frequencies up to 100,000 fold, while proofreading is expected to only increase fidelity 10 to 100 fold (80). Regardless of its actual role in the polymerization process, given its lack of specificity of mutagenesis and its location in the holoenzyme, perhaps it is the site of interaction between SOS and the replication machinery. It would be interesting indeed to screen dnaE suppressors for involvement in UV mutagenesis.

Other models have been put forth for how SOS might inhibit either proofreading or base selection (91,105). Generally, they are permutations on the ideas mentioned above concerning the various fidelity mechanisms.

Of course, SOS repair may resolve into a number of mechanisms. However, any hypothesis which purports to explain all of SOS repair must take into account the role of umuC, umuD and recA in the muta-

genic process. The genetics of DNA repair in E. coli have given us many clues concerning the mechanism of SOS repair. Hopefully the cloning of umuC and umuD genes and the identification of their gene products will now allow biochemistry to go forth and solve the puzzle of SOS repair.

#### V. Plasmid Encoded Repair Functions

Several naturally occurring plasmids isolated from a number of different bacteria display the ability to enhance the resistance of their host cells to killing by UV-irradiation (For a review see 106). Among these "UV protecting" plasmids are plasmids that also enhance the UV and chemical mutagenesis of the host cell (107,108), an observation which has led some investigators to hypothesize that some UV protecting plasmids endow their hosts with an "error prone repair" capacity. The UV protection and enhanced mutability properties appear to be linked in at least one plasmid, pKM101, because a single mutation abolishes both properties (109). Whereas in another plasmid, N3, these phenomena appear to be separable by mutation (110). As pointed out by Chernin and Mikoyan (106), there are many qualitative differences among UV-protecting plasmids with regard to their effects on UV survival and mutagenesis and there is no reason to suspect that all UV protecting plasmids act through a common mechanism.

The most widely studied UV protecting plasmid is pKM101, a deletion derivative of the clinically isolated drug resistant plasmid R46 (111,112,). Both of these plasmids were found to substantially in-

crease UV and chemical mutagenesis in their host cells. The ability to enhance mutability led to the introduction of pKM101 into the Ames Salmonella tester strains and was largely responsible for the success of this system in identifying potential carcinogens based on their bacterial mutagenicity (111,113).

In addition to increasing UV-protection and mutagenesis, pKM101 also increases Weigle-reactivation, Weigle-mutagenesis, Weigle-reactivation of single-stranded phage, and spontaneous mutagenesis (111,114,118). All of the above pKM101-mediated effects are dependent upon the recA<sup>+</sup> lexA<sup>+</sup> genotype of the host cell (111,116-118). Furthermore, pKM101 appears to enhance mutagenesis only by indirect mutagens (113).

In 1979, Walker and Dobson (114) showed that pKM101 was able to suppress the deficiencies in Weigle-reactivation, and UV and chemical mutagenesis associated with the cellular umuC mutation. This led to the hypothesis that pKM101 codes for a functional analog of the umuC gene. The region of pKM101 responsible for the suppression of umuC mutations was localized to a 2.2kb region of DNA which defines the loci termed muc (mutagenesis; UV and chemical) (119).

The mechanism through which the muc locus mediates the enhancement of UV and chemical mutagenesis remains unknown. MacPhee (120) has reported finding a DNA polymerase I-like activity in S. typhimurium polA cells containing the UV-protecting, mutagenesis-enhancing plasmid R. Utrecht. He has postulated that the repair processes

mediated by this plasmid could be explained if this polymerase was error-prone. Another UV-protecting, mutagenesis-enhancing plasmid, N3, apparently also codes for a new DNA polymerase, however this activity is still present in mutants of N3 which lack the ability to enhance mutagenesis (110). Kronish and Walker (121) found no evidence for a pKM101-coded DNA polymerase activity in extracts prepared from E. coli polA (pKM101) cells. Interestingly enough, both R. Utrecht and N3 have been found to suppress the phenotype of the umuC mutation (M. Lichtman, S. Elledge, unpublished result). Furthermore, both plasmids share DNA sequence homology with the muc region of pKM101 as demonstrated with Southern hybridization utilizing a probe specific to the muc region (M. Lichtman, S. Elledge, unpublished result). It is highly probable that these plasmids are related to one another in evolution and that they all owe their UV-protecting and mutagenesis-enhancing properties to their copy of the muc loci. An interesting difference, however, is that the UV-protecting property and mutagenesis-enhancing property of N3 have been separated by mutation while those in pKM101 have not. Perhaps there are additional genes on these plasmids which affect their mutagenesis and UV-protecting properties.

The hypothesis that pKM101 codes for a functional analog of the umuC locus has gained a great deal of support. Recently the protein products of the muc locus have been identified (122). The muc locus has actually been shown to consist of two genes, mucA, which codes for a 16kd protein, and mucB, which codes for a 45kd protein. The products of the umuC locus have also recently been identified (95, this thesis) and it has been shown to consist of two genes, umuD, which

codes for a 16kd protein, and umuC, which codes for a 45kd protein. Furthermore, both loci are arranged in an operon with identical transcriptional and regulatory patterns (this thesis). Thus it seems highly likely that pKM101 codes for analogs of the umuC and umuD genes. Their DNA sequences must have diverged significantly in evolution because they fail to cross hybridize in a Southern hybridization experiment (S. Elledge, unpublished results). It will be very interesting to see how the amino acid sequences have diverged throughout their evolutionary separation.



## Bibliography

1. Weigle, J. (1953) Proc. Natl. Acad. Sci. USA 39, 628-636.
2. Castellazzi, M., George, J., and Buttin, G. (1972) Mol. Gen. Genet. 119, 139-152.
3. Radman, M. (1974) in Molecular and Environmental Aspects of Mutagenesis (Prakash, L. et al., eds) pp. 128-142, C.C. Thomas, Publ., Springfield, Ill.
4. Defais, M., Fauquet, P., Radman, M., and Errera, M. (1971) Virology 43, 495-503.
5. Defais, M., Caillet-Fauquet, P., Fox, M.S., and Radman, M. (1976) Mol. Gen. Genet. 148, 125.
6. Witkin, E.M. (1974) Proc. Natl. Acad. Sci. USA 71, 1930-1934.
7. Witkin, E. (1976) Bacteriol. Rev. 40, 869-907.
8. Inouye, M. and Pardee, A.B., (1970) J. Biol. Chem. 245, 5813-5819.
9. Swenson, P.A. and Schemley, R.L. (1974) Int. J. Radiat. Biol. 25, 51-60.
10. Swenson, P.A., Joshi, J.G., and Schenley, R. L. (1978) Molec. Gen. Genet. 159, 125-130.
11. Pollard, E. and Randall, E.P. (1973) Radiat. Res. 55, 265-279.
12. Kogoma, T. and Lark, K.G. (1970) J. Mol. Biol. 52, 143.
13. Kogoma, F. and Lark, K.G. (1975) J. Mol. Biol. 94, 243.
14. Devoret, R. and George, J. (1974) Mutat. Res. 4, 713-734.
15. George, J., Castellazzi, M., and Buttin, G. (1975) Mol. Gen. Genet. 140, 309-332.
16. Huisman, O. and D'Ari, R. (1981) Nature 290, 797-799.
17. Radman, M. (1975) in Molecular Mechanisms for Repair of DNA (Hanawalt, P. and Setlow, R., eds) pp. 355-368, Plenum, New York.
18. Roberts, J.W. and Roberts, C.W. (1975) Proc. Natl. Acad. Sci. USA 72, 147-151.
19. Roberts, J.W., Roberts, C.W., and Craig, N.L. (1978) Proc. Natl. Acad. Sci. USA 75, 4714-4719.
20. Little, J.W., Mount, D.W., and Yanisch-Perron, C.R. (1981) Proc. Natl. Acad. Sci. USA 78, 4199-4203.
21. Brent, R. and Ptashne, M. (1981) Proc. Natl. Acad. Sci. USA 78, 4204-4208.
22. Little, J.W. and Harper, J.E. (1979) Proc. Natl. Acad. Sci. USA 76, 6147-6151.
23. Brent, R. and Ptashne, M. (1980) Proc. Natl. Acad. Sci. USA 77, 1932-1936.
24. Markham, B.E., Little, J.W., and Mount, D.W. (1981) Nuc. Acids Res., 9, 4149-4161.
25. Kenyon, C.J. and Walker, G.C. (1980) Proc. Natl. Acad. Sci. USA 77, 2819-2823.
26. Casadaban, M.J. and Cohen, S.N. (1979) Proc. Natl. Acad. Sci. USA 76, 4530-4533.
27. Little, J.W. and Mount, D.W. (1982) Cell 29, 11-22.
28. Kenyon, C. and Walker, G. (1981) Nature 289, 808-810.
29. Kenyon, C. and Walker, G. Ph.D. Thesis.
30. Fogliano, M. and Schendel, P.F. (1981) Nature 289, 196-198.

31. Bagg, A., Kenyon, C.J., and Walker, G.C. (1981) Proc. Natl. Acad. Sci. USA 78, 5749-5753.
32. Huisman, O. and d'Ari, R. (1981) Nature 290, 797-799.
33. Miller, H.I., Kirk, M., and Echols, H. (1981) Proc. Natl. Acad. Sci. USA 78, 6754-6758.
34. Tessman, E. and Peterson, P.K. (1980) J. Bacteriol. 143, 1307-1317.
35. Kenyon, C.J., Brent, R., Ptashne, M., and Walker, G.C. (1982) J. Mol. Biol. 160, 445-457.
36. Clark, A.J. and Margulies, A.P. (1965) Proc. Natl. Acad. Sci. USA 53, 541-549.
37. Howard Flanders, P. and Boyce, R.P. (1966) Radiat. Res. 6, 156-184.
38. Kato, R. and Shinoura, Y. (1977) Mol. Gen. Genet. 156, 121-131.
39. Steinborn, G. (1978) Mol. Gen. Genet. 165, 87-93.
40. Elledge, S.J. and Walker, G.C. (1983) J. Mol. Biol., 163 in press.
41. Witkin, E.M. and Wermundsen, I.E. (1979) Cold Spring Harbor Symp. Quant. Biol., 881-886.
42. Phizicky, E.M. and Roberts, J.W. (1981) Cell 25, 259-268.
43. Blanco, M., Herrera, G., Collado, P., Rebollo, J., and Botella, L. (1982) Biochemie 64, 633-636.
44. Shimizu, K. and Sekiguchi, M. (1979) Mol. Gen. Genet. 168, 37-47.
45. Sekiguchi, M., Hayakawa, H., Yamashita, K. and Shimizu, K. (1983) in New Approaches in Eukaryotic DNA Replication, in press.
46. Oeda, K., Takashi, H., and Sekiguchi, M. (1982) Nature 298, 98-100.
47. Rothman, R.H. (1978) J. Bacteriol 136, 444-448.
48. Rupp, W.D., Sancar, A., and Sancar, G.B. (1982) Biochemie 64, 595-598.
49. Rothman, R.H., Margossian, L.J., and Clark, A.J. (1979) Mol. Gen. Genet. 169, 279-287.
50. Clark, A.J., Volkert, M.R., and Margossian, L.J. (1979) in Symposia on Quantitative Biology, 43, 887-892.
51. McPartland, A., Green, L., and Echols, H. (1980) Cell 20, 731-737.
52. Clark, A.J., Volkert, M.R., Margossian, L.J., and Nagoishi, H. (1982) Mutation Res. 106, 11-26.
53. Hanawalt, P.C., Cooper, P.K., Ganesan, A.K., and Smith, C.A. (1979) Ann. Rev. Biochem. 48, 783-836.
54. Lindahl, T. (1982) Ann. Rev. Biochem. 51, 61-88.
55. Ono, J. and Shimazu, Y. (1966) Virology 29, 295-302.
56. Bleichrodt, J.F., Verheij, W.S.D. (1976) Mol. Gen. Genet. 135, 19-27.
57. Das Gupta, C.K. and Poddar, R.K. (1975) Mol. Gen. Genet. 139, 77-91.
58. Froehlich, B. (1981) Mol. Gen. Genet. 184, 416-420.
59. Mount, D.W., Walker, A.C. and Kasel, C. (1975) J. Bacteriol. 121, 1203-1207.
60. Kondo, S., Ichikawa, K., and Kato, T. (1970) Genetics 66, 187-217.

61. Radman, M., Villani, G., Boiteux, S., Kinsella, A., Glickman, B., and Spadari, S. (1977) in Origins of Human Cancer (Hiatt, H. et al., eds) pp. 903-927, Cold Spring Harbor Laboratories, Cold Spring Harbor, New York.
62. Coulondre, C. and Miller, J.H. (1977) J. Mol. Biol. 117, 577.
63. Yukiko, S., Ise, T., Kato, T., and Glickman, B. (1983) submitted to Mutation Research.
64. Kato, T., Ise, T., and Shinagawa, H. (1982) Biochemie 64, 731-733.
65. Ichikawa-Ryo, H. and Kondo, S. (1975) J. Mol. Biol. 97, 77-92.
66. Witkin, E.M. (1975) Mol. Gen. Genet. 142, 87-103.
67. Bridges, B.A. (1977) Mol. Gen. Genet. 151, 115.
68. Rathman, R.H. (1978) J. Bacteriol. 136, 444-448.
69. Coulondre, C. and Miller, J.H. (1977) J. Mol. Biol. 117, 577.
70. Foster, P.L., Eisenstadt, E., and Cairns, J. (1982) Nature 299, 265-367.
71. Foster, P.L. and Eisenstadt, E. J. Bacteriol. in press.
72. Schaaper, R.M. and Loeb, L.A. (1981) Proc. Natl. Acad. Sci. USA 78, 1773-1777.
73. Moore, P.D., Rabkin, S.D., Osborn, A.L., and Strauss, B.S. (1982) Biochemie 64, 757-762.
74. Caillet-Fauquet, P., Defois, M., and Radman, M. (1977) J. Mol. Biol. 117, 95-112.
75. Schaaper, R.M., Glickman, B.W., and Loeb, L.A. (1982) Mut. Res. 106, 1-9.
76. Miller, J. (1982) Cell 31, 5-7.
77. Topal, M.D. and Fresco, J. (1976) Nature 263, 285-289.
78. Kunkel, T.A., Shearman, C.W., and Loeb, L.A. (1981) Nature 291, 349-351.
79. Yamasaki, H., Pulkrabek, P., Grunberger, D., and Weinstein, I.B. (1977) Cancer Res. 37, 3756-3760.
80. Eisenstadt, E., Warren, A.J., Porter, J., Atkins, D., and Miller, J. (1982) Proc. Natl. Acad. Sci. USA 79, 1945-1949.
81. Foster, P.L., Eisenstadt, E., and Miller, J.H. (1983) Submitted to Proc. Natl. Acad. Sci. USA.
82. Brash, D. and Haseltine, W. (1982) Nature 298, 189-192.
83. Drake, J.W. (1969) Nature 221, 1132.
84. Fowler, R.G., Degnen, G.E., and Cox, E.C. (1974) Mol. Gen. Genet. 133, 179-191.
85. Kunkel, T.A., Eckstein, E., Mildvan, A.S., Kaplitz, R.M., and Loeb, L.A. (1981) Proc. Natl. Acad. Sci. USA 78, 6734-6738.
86. Fersht, A.R., Knill-Jones, J.W., and Tsui, W.C. (1982) J. Mol. Biol. 156, 37-51.
87. Kunkel, T.A. and Loeb, L.A. (1979) J. Biol. Chem. 254, 5718-5725.
88. Kunkel, T.A., Meyer, R.R., and Loeb, L.A. (1979) Proc. Natl. Acad. Sci. USA 76, 6331-6335.
89. Radman, M., Villani, G., Boiteux, S., Kinsella, A.R., Glickman, B.W., and Spadari, S. (1978) Cold Spring Harbor Symp. Quant. Biol. 43, 937-945.
90. Glickman, B.W. and Radman, M. (1980) Proc. Natl. Acad. Sci. USA 77, 1063-1067.

91. Bourguignon-Van Horen, F., Brotcoon, A., Caillet-Fauquet, P., Diver, W.P., Dohet, C., Doubleday, O.P., Lecomte, P., Moenhaut-Michel, G., and Radman, M. (1982) Biochemie **64**, 559-564.
92. Defais, M., Caillet-Fauquet, P., Fox, M.S., and Radman, M. (1976) Mol. Gen. Genet. **148**, 125-130.
93. Caillet-Fauquet, P., Defais, M., and Radman, M. (1977) J. Mol. Biol. **117**, 95-112.
94. Rabkin, S.D., Moore, P.D. and Strauss, B.S. (1982) Proc. Natl. Acad. Sci. USA submitted.
95. Elledge, S.J. and Walker, G.C. (1983) J. Mol. Biol., in press.
96. Radman, M., Caillet-Fauquet, P., Defais, M., and Villani, G. (1976) in Screening Tests in Chemical Carcinogenesis (Montesano et al. eds) pp. 537-545, IARC Sci. Publ. n° 12, Lyon.
97. Lackey, D., Krauss, S., and Linn, S. (1982) Proc. Natl. Acad. Sci. USA **79**, 330-334.
98. Bridges, B.A. and Mottershead, R.P. (1976) Mol. Gen. Genet. **144**, 53-58.
99. Villani, G., Boiteux, S., and Radman, M. (1978) Proc. Natl. Acad. Sci. USA **75**, 3037-3041.
100. Bridges, B.A. and Mottershead, R.P. (1978) Mol. Gen. Genet. **162**, 35-41.
101. Maurer, R. and Botstein, D. (1983) in press.
102. Degnen, G.E. and Cox, E.C. (1974) J. Bacteriol **117**, 477-487.
103. Horiuchi, T., Hisagi, M., and Sekiguchi, M. (1978) Mol. Gen. Genet. **163**, 277-283.
104. Topal, M.D. and Fresco, J.R. (1976) Nature **263**, 285-289.
105. Echols, H. (1982) Biochemie **64**, 571-576.
106. Chernin, L.S. and Mikoyan, V.S. (1981) Plasmid **6**, 119-140.
107. Arai, T. and Ando, T. (1980) Keio J. Med. **29**, 47-54.
108. Molina, A.M., Babudri, N., Tararo, M., Venturini, S., and Monti-Bragadin, C. (1979) FEMS Microbiol. Letters **5**, 33-37.
109. Walker, G.C. (1978) J. Bacteriol. **133**, 1203-1211.
110. Ando, T. and Arai, T. (1980) Keio J. Med. **29**, 55-66.
111. Mortelmans, K.E. and Stocker, B.A.P. (1979) Mol. Gen. Genet. **167**, 317-328.
112. Langer, P.J. and Walker, G.C. (1981) Mol. Gen. Genet. **182**, 268-272.
113. McCann, J., Springarn, N.E., Kobori, J., and Ames, B. (1975) Proc. Natl. Acad. Sci. USA **72**, 979-983.
114. Walker, G.C. and Dobson, P.P. (1979) Mol. Gen. Genet. **172**, 17-24.
115. Walker, G.C. (1978) J. Bacteriol. **135**, 415-421.
116. Walker, G.C. (1977) Mol. Gen. Genet. **152**, 93-103.
117. Goze, A. and Devoret, R. (1979) Mutation Res. **61**, 163-179.
118. Fowler, R.G., McGinty, L., and Mortelmans, R.E. (1979) J. Bacteriol. **140**, 929-937.
119. Shanabruch, W.G. and Walker, G.C. (1980) Mol. Gen. Genet. **179**, 289-297.
120. MacPhee, D.G. (1974) Nature **251**, 432-434.
121. Kronish, J.W. and Walker, G.C. (1979) Mutation Res. **60**, 135-142.
122. Perry, K. and Walker, G.C. (1982) Nature **300**, 278-281.

## Chapter II

### Introduction

Mutagenesis by a variety of agents such as UV, methyl methanesulfonate and 4-nitroquinoline-1-oxide is not a passive process. Rather it requires the intervention of a cellular system that processes damaged DNA in such a way that mutations result. This type of processing has been commonly referred to as "error-prone repair", a term proposed on the basis of a number of observations in Escherichia coli that suggest that the cellular events involved in producing mutations from damaged DNA are closely associated with events increasing resistance to the lethal effects of DNA damage (Radman, 1975; Witkin, 1976). However, despite the fundamental significance of such a cellular processing system to chemical and radiation mutagenesis, its biochemical mechanism has not yet been determined nor have the effects on mutagenesis and survival been shown to result from the same process.

The ability of E. coli to be mutated by UV and chemical agents can be blocked by mutations at three chromosomal loci, recA, lexA, and umuC. Mutations at the recA and lexA loci are pleiotropic (Witkin, 1976; Mount et al., 1972; Mount, 1977) and the products of these genes are involved in the regulation of a set of din (damage-inducible) genes (Kenyon and Walker, 1980, 1981; Huisman and D'Ari, 1981; Fogliano and Schendel, 1981; Brent and Ptashne, 1980, 1981; Little et al., 1981; Miller et al., 1981; Sancar et al., 1982; Little and Mount, 1982). In contrast, mutations at the umuC locus (Kato and Shinoura,

1977; Steinborn, 1978) make cells nonmutable, slightly UV-sensitive, and deficient in Weigle-reactivation of damaged bacteriophage but do not affect the induction of other SOS responses. Thus the umuC gene product(s) is the best candidate for a protein(s) that plays a key mechanistic role in the process of error-prone repair. By obtaining a umuC-lac fusion we recently have been able to demonstrate that expression of the umuC gene is induced by DNA damage and to analyze its regulation by the recA<sup>+</sup> and lexA<sup>+</sup> gene products (Bagg *et al.*, 1981).

In order to facilitate a systematic analysis of the cellular functions required for UV and chemical mutagenesis we decided to clone umuC. In this paper we report the successful cloning of the umuC locus of E. coli, demonstrate that it consists of two genes, identify the protein products of these genes, and analyze their transcriptional organization and regulation.

## Results

### Isolation of a umuC::Tn5 Mutation

We initially attempted to isolate the umuC gene by cloning random fragments of E. coli DNA into the vector pBR322 and screening these hybrid plasmids for their ability to complement the UV nonmutability of a umuC mutant. When this approach proved unsuccessful, we then decided to employ a strategy of screening for the umuC gene by hybridization using a probe to the umuC region generated with the aid of an insertion mutation in the umuC locus.

The umuC122::Tn5 mutant was isolated by using a localized mutagenesis procedure that took advantage of the fact that umuC is approximately 50% linked to purB by P1 transduction (Kato and Shinoura, 1977). First we obtained approximately 20,000 independent derivatives of a umuC<sup>+</sup> strain (AB1157) that had the transposon Tn5 inserted randomly into the bacterial chromosome; Tn5 encodes resistance to kanamycin (Km). These Tn5 derivatives were pooled into four sets of 5,000 transductants each and P1 bacteriophage were grown on these pools. These P1 lysates were then used to transduce a purB strain to Pur<sup>+</sup> Km<sup>R</sup>. Out of 128 such Pur<sup>+</sup> Km<sup>R</sup> derivatives, one (GW2100) did not display increased reversion of its his-4 ochre mutation or its argE3 ochre mutation after exposure to UV irradiation (15 J/m<sup>2</sup>) or to methyl methanesulfonate and, furthermore, was slightly UV-sensitive. Thus this derivative had the same phenotypic characteristics as previously isolated umuC mutations (Kato and

Shinoura, 1977; Steinborn, 1978; Bagg et al., 1981). When  $\phi$  bacteriophage grown on GW2100 were used to transduce a purB strain to  $\text{PurB}^+$  both the  $\text{Km}^R$  and the nonmutability characteristic cotransduced and were 40% linked (26/65) to the purB locus. When this same  $\phi$  lysate was used to transduce a umuC121::Mud(Ap, lac) strain (GW1103) (Bagg et al., 1981) to  $\text{Km}^R$ , all the transductants tested (41/41) were  $\text{Ap}^S$  and temperature resistant indicating loss of the Mud(Ap, lac) phage, but remained nonmutable. Thus the position of the Tn5 insertion in GW2100 is closely linked to the position of the Mud(Ap, lac) insertion in GW1103. Moreover, the introduction the the mutagenesis-enhancing plasmid, pKM101, into GW2100 suppressed the nonmutability and UV-sensitivity of the strain as it does with known umuC mutants (Walker and Dobson, 1979; Bagg et al., 1981). On the basis of these observations, we concluded that GW2100 contained an insertion of Tn5 at the umuC locus.

#### Isolation of a umuC-Specific Probe

The isolation of the umuC122::Tn5 mutant (GW2100) provided us with a convenient way of obtaining a specific probe for umuC. Southern blotting analysis of GW2100 chromosomal DNA using a Tn5-specific probe revealed that BamHI cleavage generated a 3.2 kb fragment containing the neomycin phosphotransferase gene of Tn5, one of the inverted repeats of Tn5, and approximately 0.2 kb of the chromosomal DNA adjacent to the position of the Tn5 insertion. This 3.2 kb BamHI fragment was cloned by ligating BamHI-cleaved GW2100 DNA with BamHI-cleaved pBR322 and transforming AB1157, selecting for



Ap<sup>R</sup>Km<sup>R</sup>. A hybrid plasmid carrying this 3.2 kb BamHI fragment was called pSE100.

In order to prove that the clone we had obtained contained DNA sequences homologous to the umuC region, we isolated the 3.2 kb BamHI restriction fragment, nick-translated it, and used it to probe AB1157 (umuC<sup>+</sup>), GW2100 (umuC122::Tn5), and GW1103 [umuC121::Mud(Ap, lac)] DNA cut with KpnI. KpnI was chosen because it fails to cut within Tn5 (Jorgensen *et al.*, 1979) or Mud(Ap, lac) (M. O'Connor, personal communication). As shown by the resulting Southern blot (Fig. 2-1A), the 3.8 kb KpnI band of AB1157, which was detected by the probe, was altered to give KpnI fragments of 9.5 kb in GW2100 and >35 kb in GW1103 as would be expected for the insertion of Tn5 (5.4 kb) and Mud(Ap, lac) (37 kb) respectively. This experiment demonstrated that the 3.2 kb probe would identify a region of E. coli DNA shown by genetic studies to be associated with the umuC locus.

We then used this nick-translated umuC-specific probe to screen a Charon 28  $\lambda$  library of E. coli DNA for bacteriophage bearing sequences homologous to the probe. By screening 5000 plaques we were able to identify two  $\lambda$  clones carrying such DNA.

#### Complementation by $\lambda$ -umuc<sup>+</sup>

Since Charon 28 clones lack cI and the genes needed for lysogenization and have no selectable marker, there was no easy way of obtaining lysogens in order to test for complementation of umuC

Figure 2-1. Identification of a umuC Probe and Analysis of Insertions In and Near the umuC Locus.

Chromosomal DNA from the strains indicated below were cleaved with the indicated restriction enzymes, fractionated on 0.8% agarose gels and blotted to nitrocellulose as described in Experimental Procedures. Probes employed were the BamHI fragment from pSE100 containing both Tn5 DNA and chromosomal DNA from the umuC locus for (A) and the 0.95 kb BglIII fragment illustrated in Figure 3 for (B) and (C).

(A) Lane 1, GW2100 (umuC122::Tn5) DNA; lane 2, AB1157 (umuC<sup>+</sup>) DNA; lane 3, GW1103 [umuC121::Mud(Ap, lac)] DNA; all were cleaved with KpnI. The lighter 9.6 kb band in lane 2 is a contaminant from lane 1 since it was absent in other AB1157 KpnI digests.

(B) Lanes 1, 3 and 5 are AB1157 (umuC<sup>+</sup>) DNA; lanes 2, 4, and 6 are GW1003 (Tn5 between purB<sup>+</sup> and umuC) DNA. Restriction enzymes used were PvuI for lanes 1 and 2, BamHI for lanes 3 and 4, and EcoRI for lanes 5 and 6.

(C) Lanes 1, 3, and 5 are AB1157 (umuC<sup>+</sup>) DNA; lanes 2, 4, and 6 are GW1103 [umuC121::Mud(Ap, lac)] DNA. Lanes 1 and 2 are cleaved with PstI; lanes 3 and 4 are cleaved with EcoRI; lanes 5 and 6 are cleaved with both PstI and EcoRI.

**A**

**1**

**2**

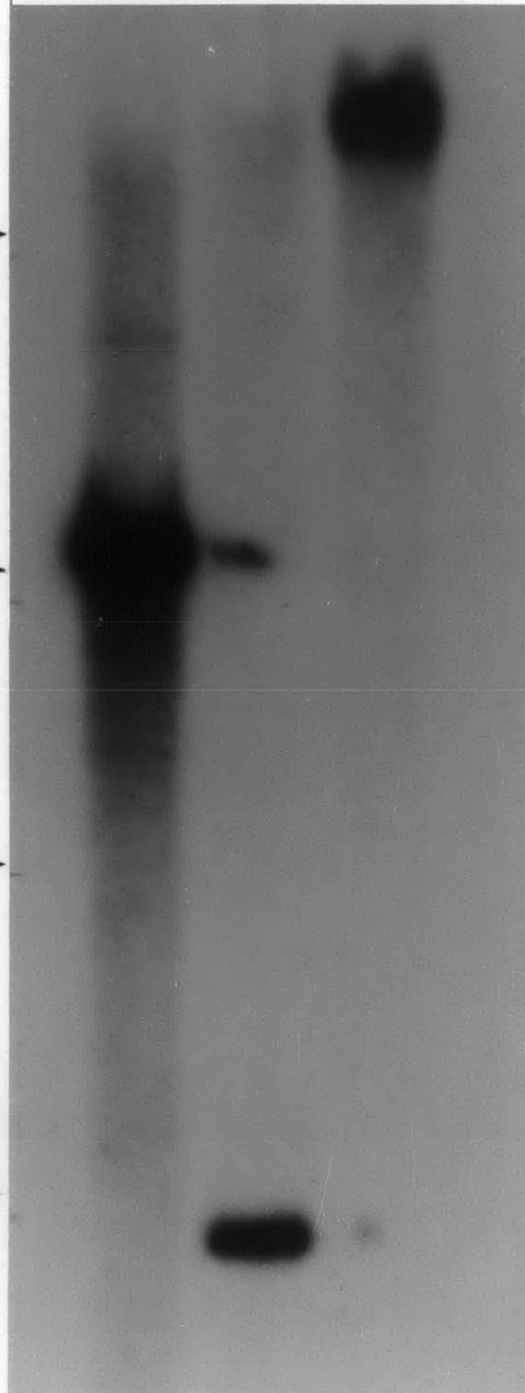
**3**

**23.3** →

**9.5** →

**6.4** →

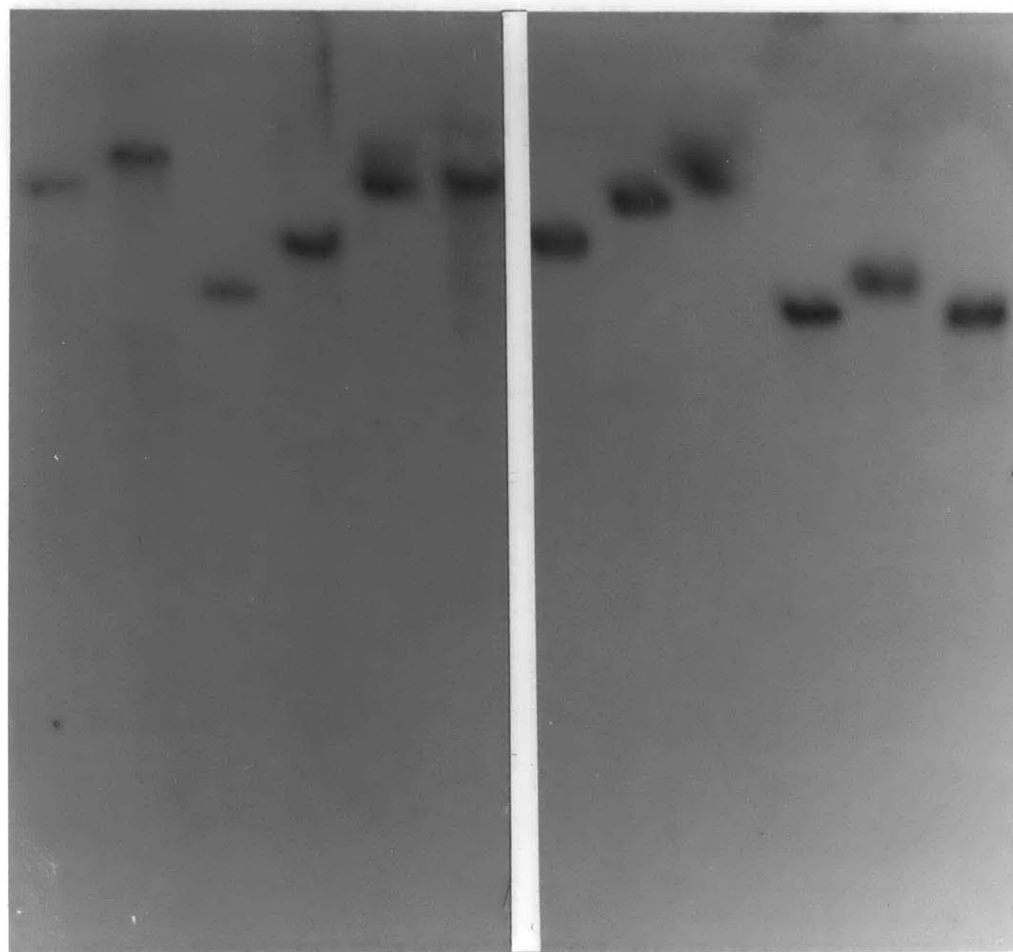
**4.2** →



**B**

**C**

**1 2 3 4 5 6 1 2 3 4 5 6**



— 23.3

— 9.5

— 6.4

— 4.2

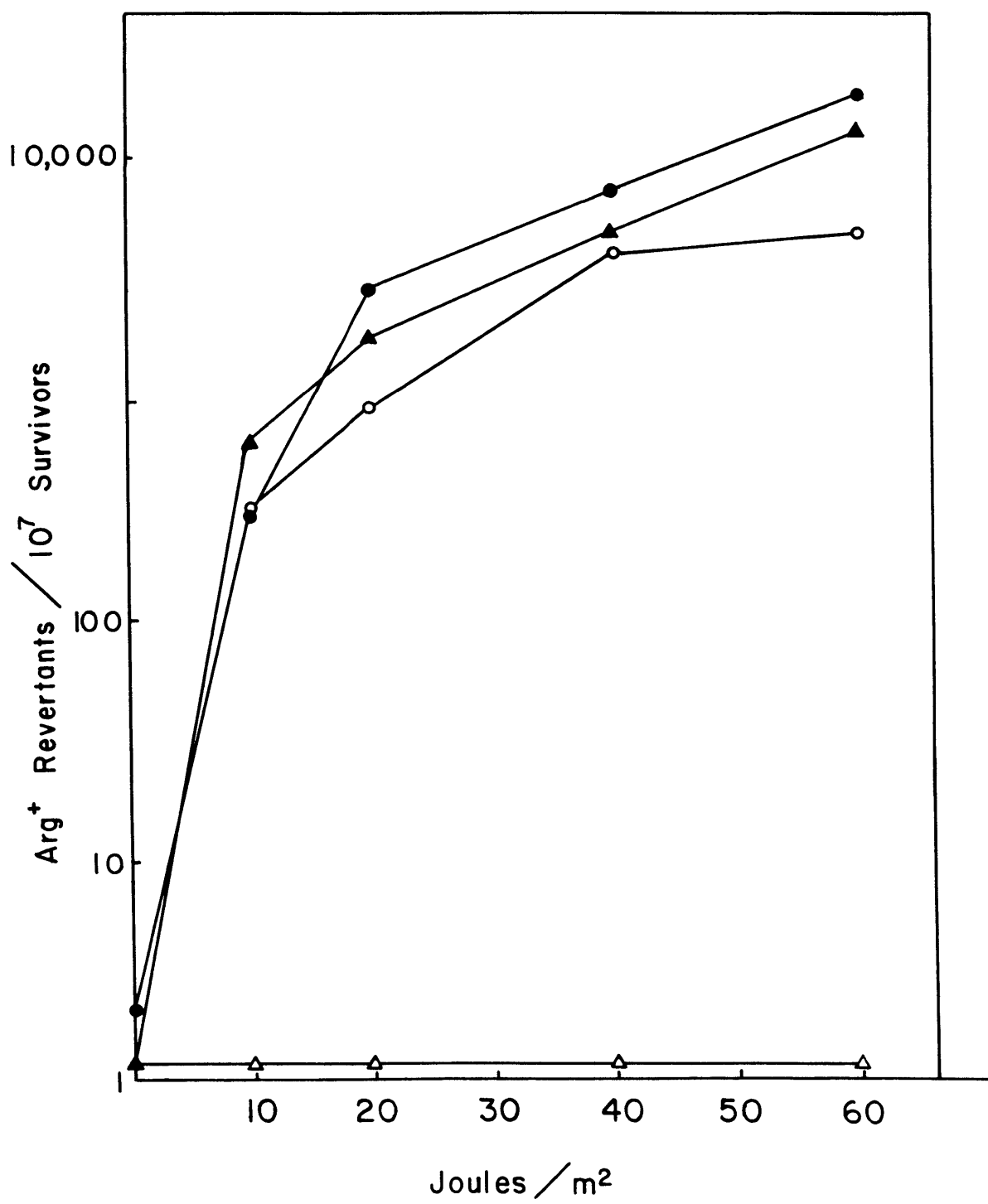
mutations by the bacteriophage carrying umuC-specific DNA. However it seemed likely that, if cells were temporarily supplied with umuC function and then were damaged with UV light, the mutational process could be completed even though the umuC gene was not stably inherited. Thus we introduced pKB280, a plasmid causing overproduction of the  $\lambda$  repressor (Backman and Ptashne, 1978), into GW2100 (umuC122::Tn5) in order to prevent expression of bacteriophage genes. The  $\lambda$  bacteriophage derivatives carrying umuC-specific DNA were adsorbed to exponentially growing GW2100(pKB280) cells. The cells were then UV-irradiated and examined for the reversion of their his-4 ochre mutation. Cells infected with either of these  $\lambda$  derivatives gave approximately 1000 fold increases in reversion frequency compared to cells infected with  $\lambda$  derivatives not homologous to our umuC-specific probe or to mock-infected cells (Fig. 2-2). Thus we concluded that both these  $\lambda$  derivatives coded for umuC<sup>+</sup> function and continued to study the one termed  $\lambda_{SE14}$ . The amount of mutagenesis observed in these experiments [umuC<sup>+</sup> +  $\lambda_{SE14}$ ) > (umuC +  $\lambda_{SE14}$ ) > umuC<sup>+</sup> >> umuC)] correlated with expected umuC<sup>+</sup> gene dosage, suggesting that umuC function is normally rate-limiting for UV mutagenesis in an AB1157 background.

#### Subcloning the umuC Region

In order to facilitate further studies, we wished to clone the umuC gene into a plasmid vector. Since we had had difficulty identifying the umuC gene on the high copy number vector pBR322 we constructed a lower copy number plasmid suitable for cloning Sau3A1

Figure 2-2. Identification of  $\lambda$  umuC<sup>+</sup> Clones by Complementation in Transient Merodiploids.

Cells containing pKB280 (a multicopy plasmid carrying the  $\lambda$ cI<sup>+</sup> gene) were infected with  $\lambda$ <sub>SE14</sub>, a clone which appeared positive in the  $\lambda$  plaque hybridization screen, at a m.o.i. of 2.5. After a 15 min period for phage adsorption, the cells were irradiated with various doses of UV light and then plated for Arg<sup>+</sup> revertants and survival as described in Experimental Procedures. GW2100(pKB280) mock infected ( $\Delta$ ); AB1157(pKB280) mock infected (O); GW2100(pKB280) infected with  $\lambda$ <sub>SE14</sub> ( $\blacktriangle$ ); AB1157(pKB280) infected with  $\lambda$ <sub>SE14</sub> ( $\bullet$ ). GW2100(pKB280) infected with wild type  $\lambda$  or  $\lambda$  clones not carrying DNA homologous to umuC behaved essentially as GW2100(pKB280) mock infected cells (data not shown). Note that Arg<sup>+</sup> revertants/ $10^7$  survivors is plotted on a logarithmic scale.



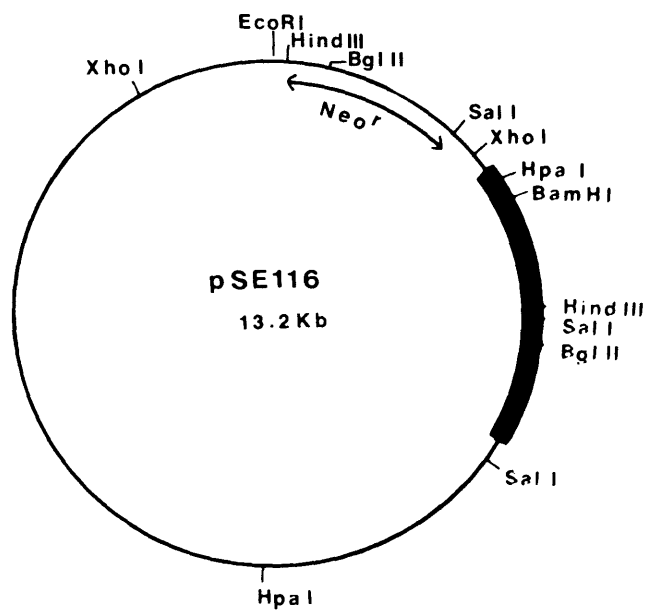
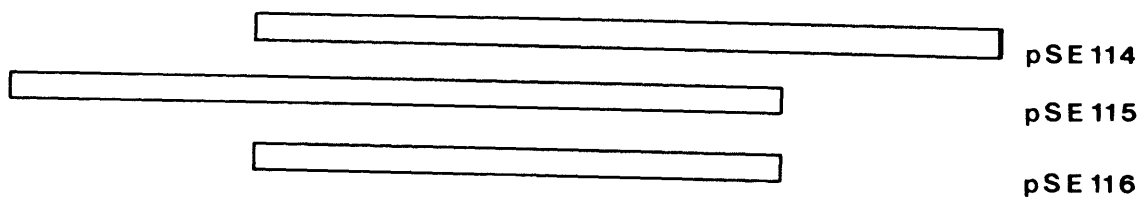
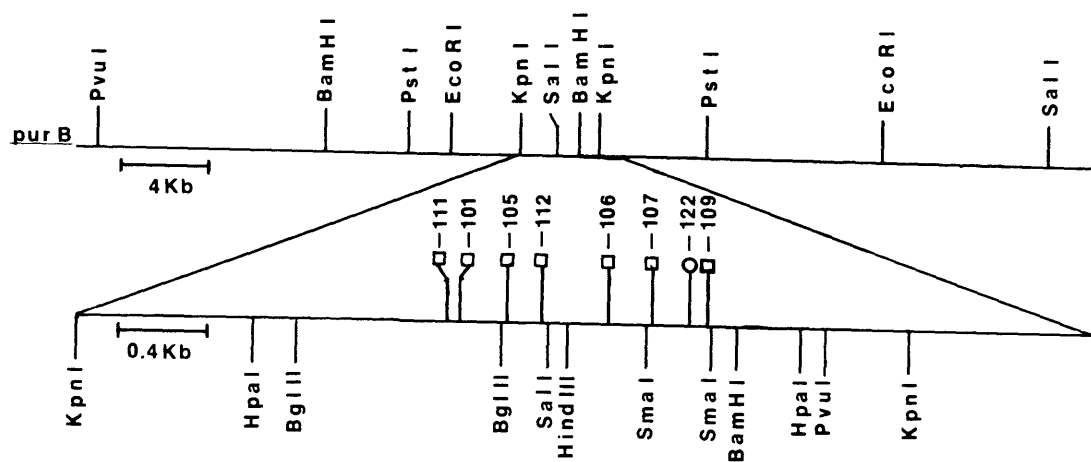
fragments. The HindIII-BamHI fragment of the transposon Tn5 which carries the neomycin phosphotransferase gene was subcloned from pSE100 into HindIII, BamHI-cleaved pSC101 (Cohen and Chang, 1977) DNA thus replacing  $Tc^R$  with  $Km^R$  and introducing a BamHI site outside of the drug resistance gene. This plasmid was designated pSE101.

DNA from  $\lambda_{SE14}$  was partially digested with Sau3A1 and 6-9 kb fragments were isolated. These fragments were cloned into the BamHI site of pSE101 and transformed into a umuC36 mutant (TK610). The transformants were then screened for their ability to show increased reversion of their his-4 mutation after UV-irradiation. Four plasmids, pSE110-pSE113 were chosen that complemented the nonmutability of a umuC36 mutant. Since pSE110 and pSE111 contained approximately the same fragment but in opposite orientations, it seemed likely that the fragment carried the umuC promoter. Analysis of these clones narrowed the possible location of umuC to a 4 kb region. A second subcloning into pSE101 was then carried out using fragments in the 2-5 kb range generated by partial Sau3A1 digestion of  $\lambda_{SE14}$ . This yielded the plasmids pSE114 and pSE115 whose 4 kb inserts are illustrated in Fig. 2-3. In order to make a still smaller derivative, we took advantage of the fact that the inserted DNA in pSE114 and pSE115 was in the same orientation on the vector. By replacing the smaller of the two HindIII fragments of pSE114 with that of pSE115, we generated pSE116, which contained an insert of 2.4 kb (Fig. 2-3).



Figure 2-3. The Restriction Map of the umuC Region of E. coli and the Location of umuC::Tn1000 Insertions.

The upper line represents a partial restriction map of the umuC region of E. coli showing DNA from ca. 25 to 26 minutes on the E. coli genetic map. This map was generated using Southern blotting analysis with the probe generated from pSE100 described above. The data from the Southern blotting analysis shown in figure 1(B) allowed us to orient the restriction map, placing purB on the left. The second line is a more detailed map of the umuC locus generated by restriction endonuclease cleavage of cloned DNAs and analysis by agarose gel electrophoresis. ( ☐ ) represents the location of umuC::Tn1000 insertions isolated on pSE110 as described in the Experimental Procedures. (O) represents the location of the chromosomal umuC::Tn5 insertion in GW2100. The solid bars beneath this map indicate the regions of DNA subcloned into the BamH1 site of pSE101 using Sau3A1-generated partials of  $\lambda_{SE14}$ . A restriction map of pSE116, the smallest umuC derivative of pSE101 constructed, is shown at the bottom of the figure.



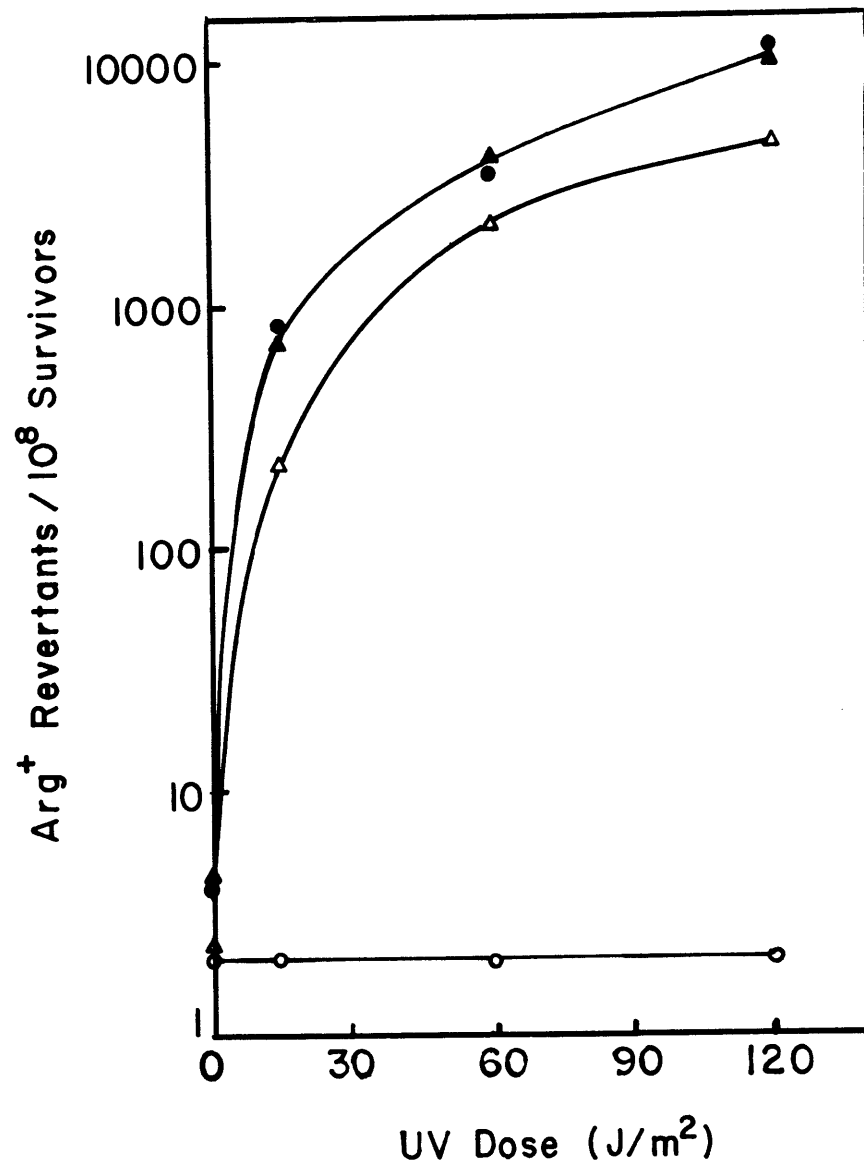
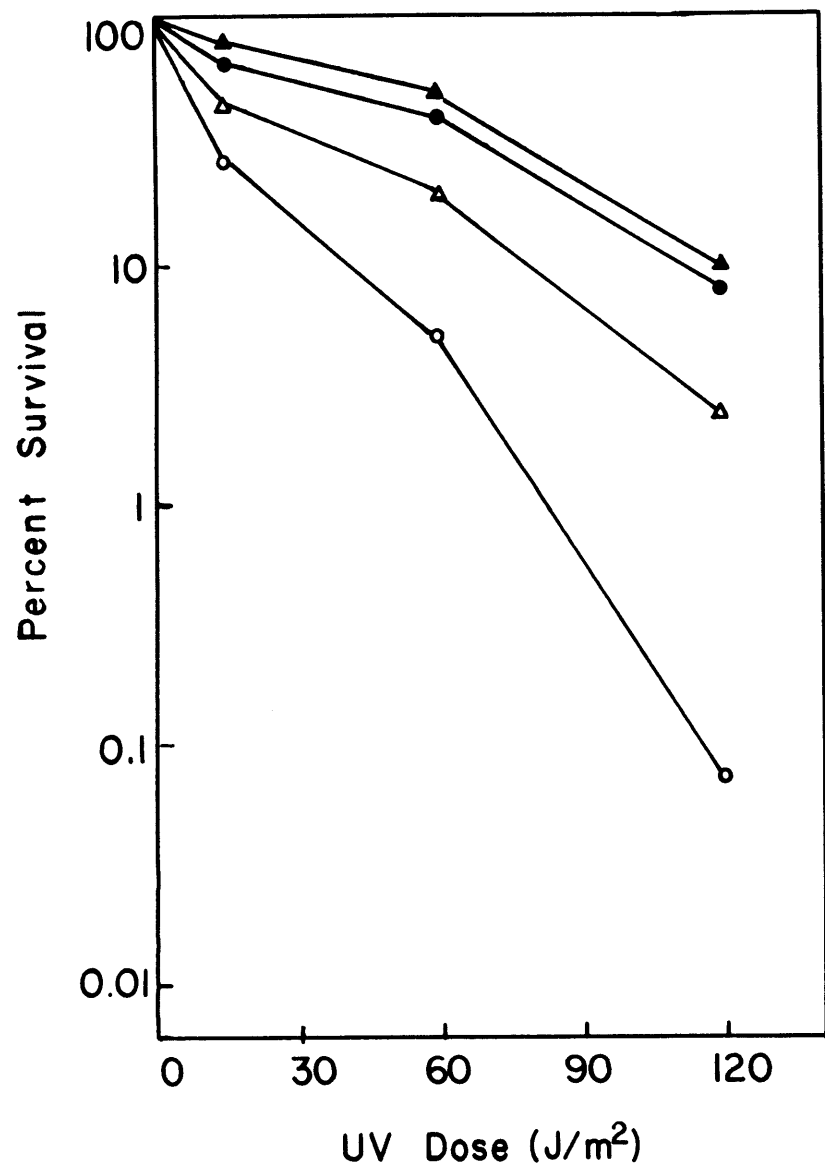
We wished to examine the ability of pSE116 to complement GW2100 (umuC122::Tn5), but both the plasmid and the strain coded for Km<sup>R</sup>. We therefore replaced the small EcoRI-PstI fragment of pSE116 with the EcoRI-PstI fragment of pDPT427 (Sninsky et al., 1981) that coded for spectinomycin resistance, thus generating pSE137. pSE137 was introduced into GW2100 and UV induced Arg<sup>+</sup> reversions were scored by plating on medium containing low levels of arginine thus allowing both mutagenesis and survival to be measured under identical conditions. This plasmid and its parent, pSE116, must carry the umuC<sup>+</sup> gene since introduction of pSE137 into GW2100 (umuC122::Tn5) restored the ability of the strain to be mutagenized by UV and made it more resistant to killing by UV (Fig. 2-4). Once again it seems that the umuC gene product may be rate limiting for UV mutagenesis in a normal cell since the introduction of multiple copies of the umuC gene increased both mutability and UV-resistance above that of a wild type cell.

#### Identification of Proteins Required for UV Mutagenesis

We then obtained insertions of the transposon Tn1000 ( $\gamma\delta$ ) (Sancar and Rupp, 1979) in pSE110 and screened these in a umuC36 background (TK610) for derivatives that failed to restore the UV-mutability of the cells by assaying for reversion to His<sup>+</sup>. As shown in Fig. 2-3, the Tn1000 insertions in the seven derivatives we examined mapped within a 1.3 kb region of DNA that must be required for mutagenesis. Results obtained with other mutations discussed below indicate that this required region is at least 1.7 kb in length.

Figure 2-4. UV Mutagenesis and Survival in a umuC Strain Carrying a umuC<sup>+</sup> Plasmid.

Cells were grown in minimal media to a density of ca.  $1 \times 10^8$ , spun down and resuspended in 0.85% saline. They were then placed in glass petri dishes and irradiated with the indicated UV dose. Mutagenesis and survival was measured as described in the Experimental Procedures. (▲) AB1157(pSE137); (●) GW2100(pSE137); (Δ) AB1157; (○) GW2100.



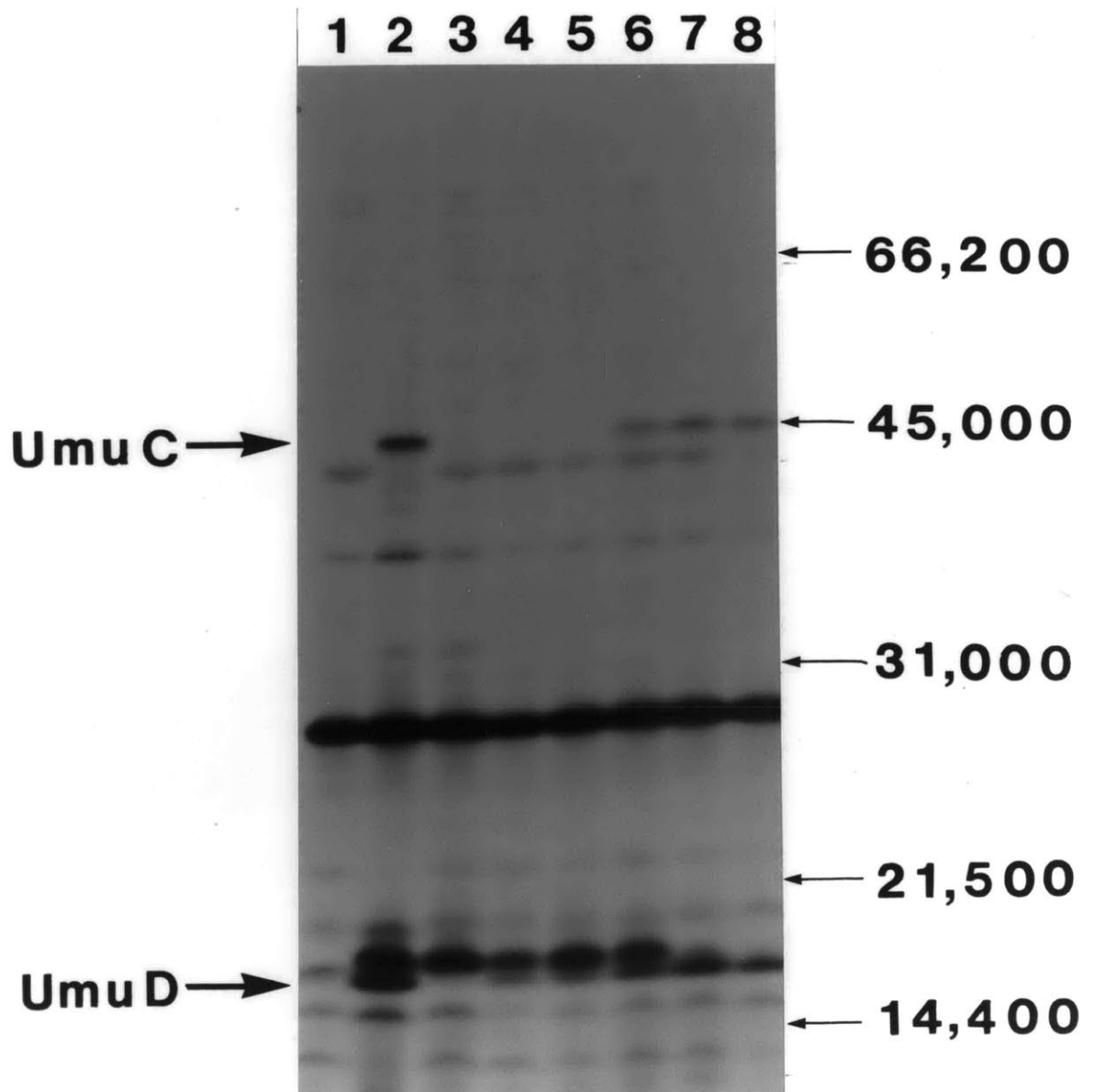
Furthermore, we were able to orient our restriction map in Fig. 2-3 with respect to the purB gene by probing chromosomal DNA from a strain with a Tn5 insertion that maps genetically between purB and umuC. The data is shown in Fig. 2-1B. The Tn5 insertion alters bands in the PvuI and BamHI digests but not in the EcoRI digest indicating that the Tn5 is located between the EcoRI and BamHI sites to the left of the umuC region as shown in Fig. 2-3. Therefore the purB gene also maps to the left of the umuC gene.

Several of the Tn1000 insertion mutants of pSE110 were introduced into RB901, a spr-51 ΔrecA21 strain and plasmid-coded proteins were examined by the maxicell technique (Sancar *et al.*, 1982). This strain was chosen because we have previously shown that umuC is negatively regulated by the lexA<sup>+</sup> product and that introduction of a spr mutation (a null allele of the lexA gene) leads to constitutive expression of the umuC gene even in a recA background (Bagg *et al.*, 1981). As shown in Fig. 2-5, pSE116 coded for two proteins, one of 45 kilodaltons and one of 16 kilodaltons. Two Tn1000 insertions in pSE110, which are 1.2 kb apart, eliminate the 45 kilodalton protein. However, another Tn1000 insertion, the furthestmost to the left in Fig. 2-3, eliminated both the 45 kilodalton protein and the 16 kilodalton protein. Since Tn1000 insertions are known to be polar (Broker, 1977) this suggests that the two proteins are in the same operon with the 16 kilodalton protein being upstream of the 45 kilodalton protein.

Other proteins on pSE110 not involved in mutagenesis were mapped by virtue of their absence in pSE114 or pSE115. Thus an 18 kilodalton

Figure 2-5. Identification of the Proteins Encoded by the umuC Locus of E. coli.

Plasmids containing various subclones and umuC::Tn1000 insertions as illustrated in Figure 3 were introduced into RB901(recA spr51). Plasmid-encoded proteins were labeled by the maxicell method of Sancar et al (1981) and analyzed on 10 to 20% gradient SDS-polyacrylamide gels as described in the Experimental Procedures. Lane 1, pSE101; lane 2, pSE110; lane 3, pSE110 umuC111::Tn1000; lane 4, pSE110 umuC101::Tn1000; lane 5, pSE110 umuC109::Tn1000; lane 6, pSE114; lane 7, pSE115; lane 8, pSE116. The 34 kilodalton protein which appears to be absent in the Tn1000 insertions appears on longer exposures. That protein is deleted in pSE114 and pSE116.





protein maps to the right of the mutagenesis genes and a 34 kilodalton protein maps to the left (Fig. 2-5).

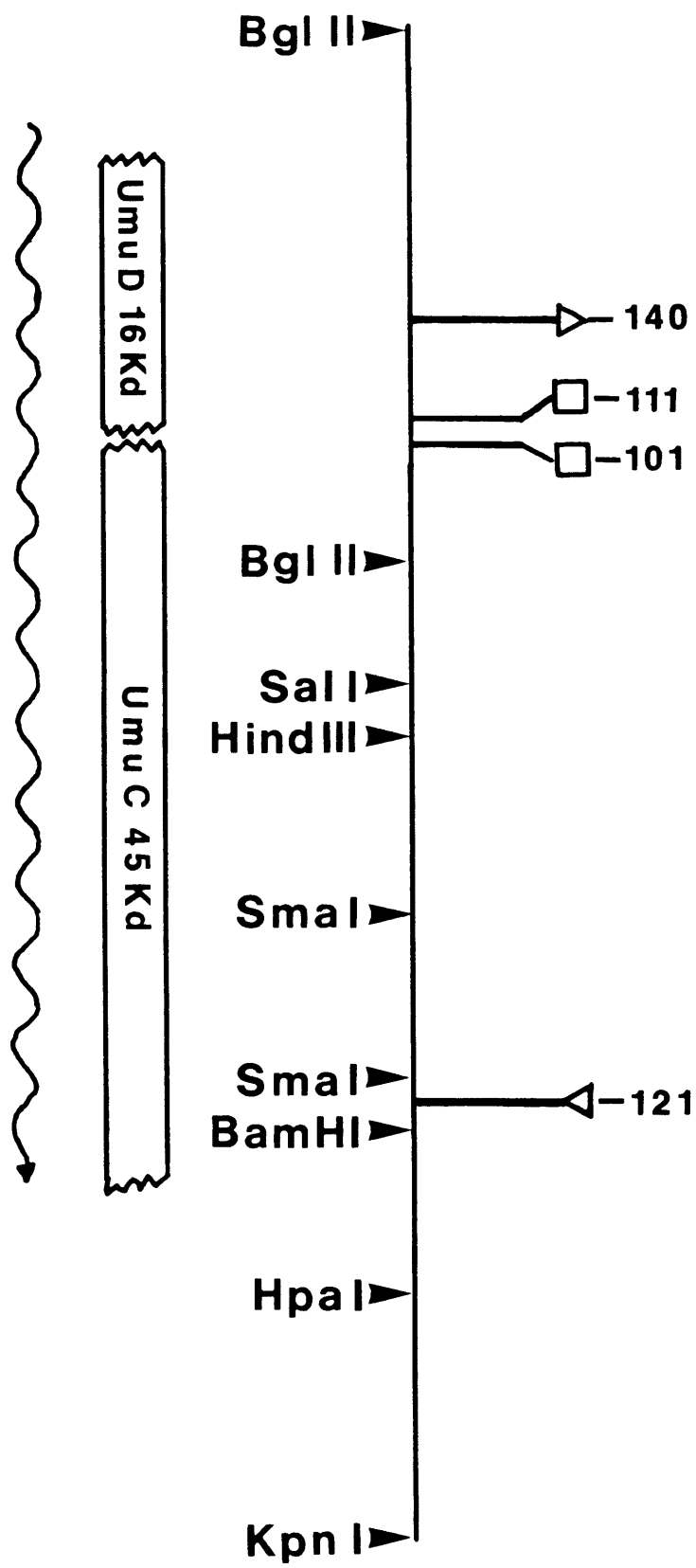
#### Direction of Transcription and Regulation

Southern blotting analysis of the previously isolated umuC121::Mud(Ap, lac) insertion placed that insertion mutation within the DNA coding for the 45 kilodalton protein as illustrated in Fig. 2-6. We have used this information and the restriction map of Mud(Ap, lac) (M. O'Connor, personal communication) to determine the direction of transcription of the 45 kilodalton protein (Fig. 2-10). Given that the half of the Mud(Ap, lac) phage DNA containing the lac genes has an EcoRI restriction site at 4.5 kb from the end and a PstI site at 10 kb from the end, while the immunity half has a PstI site at 1.5 kb and an EcoRI site at 4.0 kb from the end, the fact that the size of the EcoRI fragment of GW1103 is unchanged in the EcoRI-PstI double digest using a probe to the left of the insertion (Fig. 2-10), indicates that the lac end of Mud(Ap, lac) is nearest the probe. Therefore transcription of the 45 kilodalton protein is from left to right as shown in Fig. 2-6.

To examine the regulation of the small protein, an operon fusion was constructed by using in vitro techniques to introduce a piece of DNA carrying the  $\beta$ -galactosidase structural gene but no promoter into the region of pSE110 coding for the small protein. The position of the inserted lacZ gene in the resulting plasmid, pSE140, is shown in Fig. 2-6; restriction mapping of the fusion indicated that

Figure 2-6. The Position and Transcriptional Organization of the umuC and umuD Gene Products.

The upper line represents an expanded restriction map of the umuC region. Insertions in this region are indicated by the following symbols: ( $\Delta$ ) represents the insertion in pSE140, a lac operon fusion constructed using pSE110 and in vitro methods described in the Experimental Procedures; (O) represents Tn1000 insertions in pSE110; ( $\nabla$ ) represents the umuC::Mud(Ap, lac) insertion in GW1103. The lower blocks represent the region of the restriction map thought to code for the two polypeptides, umuC and umuD, observed in Figure 5. The inner boundaries of these proteins were based on the two Tn1000 insertions shown. umuC111::Tn1000 proved to be umuC-umuD with respect to maxicell protein analysis while umuC101::Tn1000 proved to be umuC-umuD and thus umuC111::Tn1000 marked the end of umuD polypeptide and umuC101::Tn1000 marked the beginning of the umuC polypeptide. The size of these regions were calculated based on the coding capacity needed to produce the umuD 16 kilodalton polypeptide and the umuC 45 kilodalton polypeptide. The lower arrow indicates the direction of transcription as determined from the Southern blotting analysis of the umuC121::Mud(Ap, lac) insertion shown in Figure 1C and the orientation of the lac operon insertion in pSE140 (data not shown).



transcription of the 16 kilodalton protein must be from left to right. In cells containing pSE140,  $\beta$ -galactosidase was induced 14 fold by thermal induction in a tif background and was also induced by UV-irradiation. The observation that the small protein was transcribed in the same direction as the 45 kilodalton protein and had similar regulatory characteristics strengthened the suggestion that the two proteins were in the same operon. The above results did not rule out the formal possibility that the 16 kilodalton protein was a transcriptional activator for synthesis of the 45 kilodalton protein. However we consider this latter possibility unlikely since introduction of the plasmid pSE140 into our umuC121::Mud(Ap, lac) strain did not result in restoration of UV-mutability. In that strain, the chromosome expresses the 16 kilodalton protein but has an insertion in the gene for the 45 kilodalton protein, while the plasmid has an insertion in the 16 kilodalton protein and carries an intact but unexpressed copy of the 45 kilodalton protein. Complementation would have been expected if the 16 kilodalton protein functioned as a transcriptional activator of the 45 kilodalton protein but not if the genes for the two proteins were in the same operon. Furthermore, by subcloning from pSE140, we were able to localize the region containing the operator and promoter of umuC to the 500 bp of DNA between the BglIII site and the lac insertion as shown in Fig. 2-6.

Our analysis of the regulation of the umuC gene that was carried out using the Mud(Ap, lac) fusion suggested that the lexA product functioned as the direct repressor of the umuC gene (Bagg et al., 1981) and we were interested in seeing whether evidence for such

regulation could be observed in maxicells. Thus we examined proteins synthesized from the plasmid pSE116 in maxicells prepared from either a recA strain or a recA spr strain but could observe no quantitative difference in the amount of umuC synthesized relative to control proteins. Since it seemed possible that the levels of lexA protein in maxicells might be quite low, we introduced the plasmid pSE152, a spectinomycin-resistant derivative of pRB160 (Brent and Ptashne, 1980) which carries the lexA<sup>+</sup> gene, into the pSE116-bearing strains. An analysis of the protein products synthesized in maxicells derived from these strains is shown in Fig. 2-7. The 16 and 45 kilodalton proteins coded for by pSE116 were synthesized at very low levels when the lexA<sup>+</sup> plasmid was present. These observations are consistent with the results discussed above indicating that the two proteins are coordinately regulated and strengthens our conclusion that lexA acts as the repressor of their respective genes.

#### Identification of Two Complementation Groups and Reassignment of Previously Described umuC Alleles

Analysis of the chromosomal insertions umuC121::Mud(Ap, lac) and umuC122::Tn5 indicated that the 45 kilodalton protein was essential for mutagenesis. Since the 16 kilodalton protein resides in the same operon, it seemed likely to also play a role in mutagenesis. In order to investigate this possibility, we introduced the pSE110 umuC101::Tn1000 plasmid which expresses only the 16 kilodalton protein (Fig. 2-5) into (AB1157) backgrounds containing the umuC36 and umuC44 alleles. The data illustrated in Fig. 2-8 shows that the presence of the 16 kilodalton protein completely restores the ability to carry out

Figure 2-7. Repression of the Synthesis of umuC and umuD Protein Synthesis by the lexA Protein in Maxicells.

Plasmid encoded proteins of pSE116 were analyzed for the synthesis of the umuC and umuD gene products in cellular backgrounds differing in the amount of lexA protein produced. Proteins were labeled by the maxicell method and analyzed on 10%-20% SDS polyacrylamide gels as described previously. Lane 1, RB901(pSE101); lane 2, SY822 (pSE116); lane 3, RB901(pSE116); lane 4, RB901(pSE116)(pSE151); lane 5, RB901 (pSE116)(pSE152); lane 6, RB901 (pSE152).

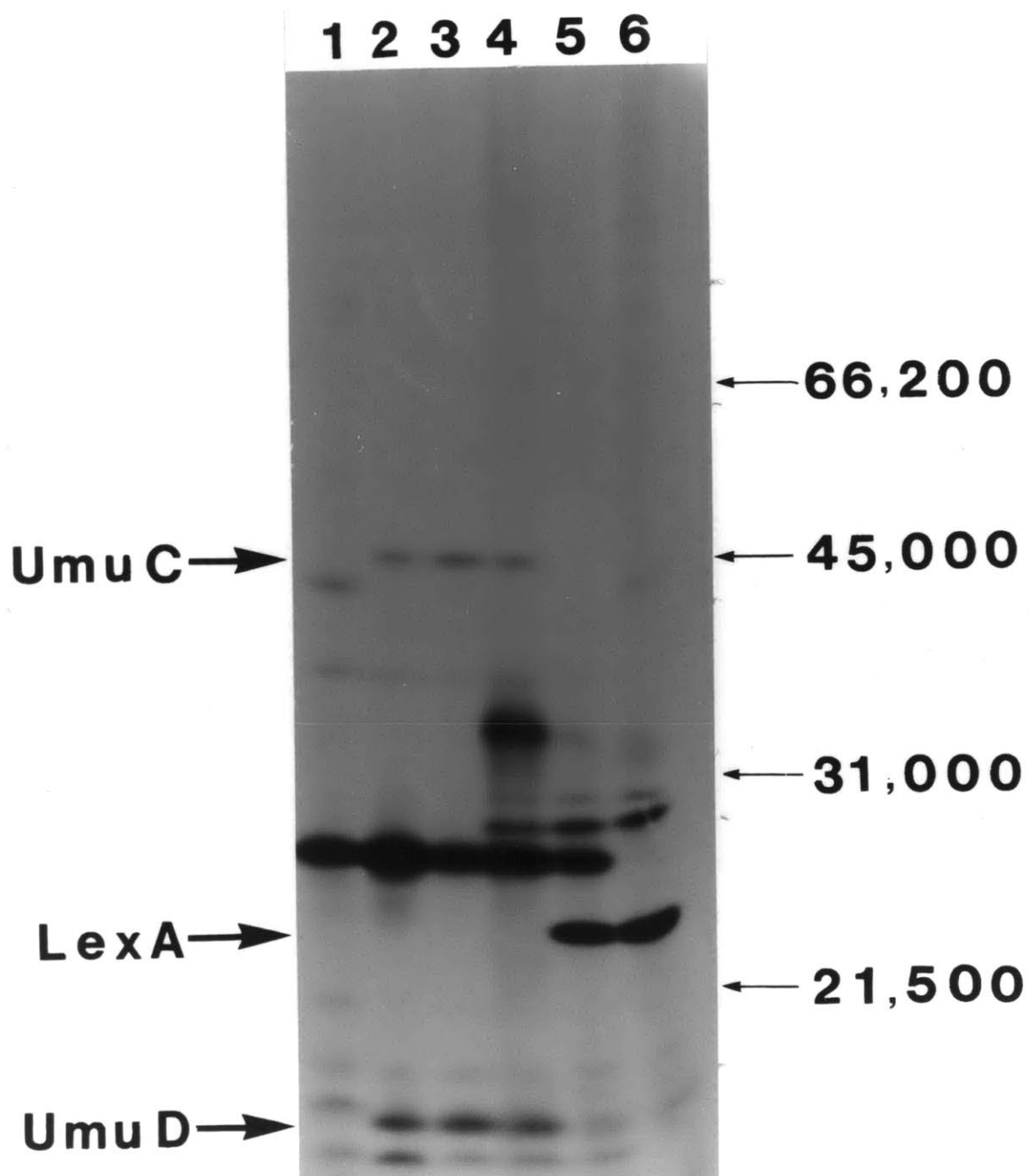
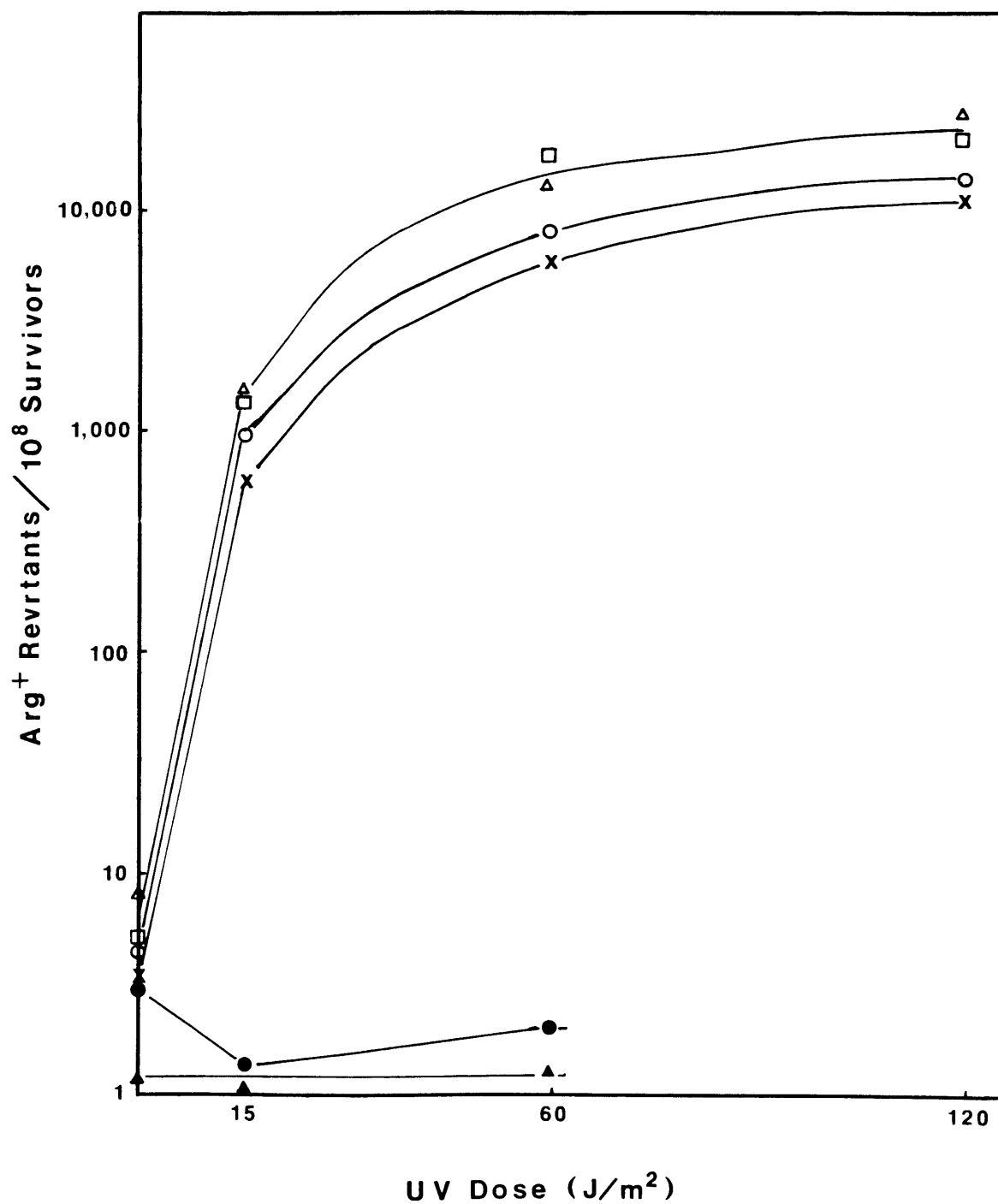


Figure 2-8. Complementation of UV-mutagenesis and Survival of umuC and umuD Mutants.

Strains were grown to a density of  $2 \times 10^8$ , centrifuged and resuspended in 0.85% saline, and transferred to glass petri dishes. Cells were UV irradiated and then plated for mutagenesis and survival as described in the Experimental Procedures.

(x) AB1157; ( $\Delta$ ) AB1157 umuC36 (pSE110); (O) AB1157 umuC44 (pSE110); ( $\square$ ) AB1157 umuC44 (pSE110 umuC101::Tn1000); ( $\blacktriangle$ ) AB1157 umuC44 (pSE110 umuC111::Tn1000); ( $\bullet$ ) AB1157 umuC36, (pSE110 umuC101::Tn1000). AB1157 umuC36 (pSE110 umuC111::Tn1000), AB1157 umuC36 and AB1157 umuC44 were essentially the same as ( $\blacktriangle$ ) and ( $\bullet$ ) (data not shown).





UV mutagenesis to strains bearing the umuC44 allele while failing to complement the umuC36 allele. Furthermore, pSE110 umuC111::Tn1000 which has an insertion in the gene for the 16 kilodalton protein fails to complement either umuC allele; this is consistent with its inability to synthesize either protein in maxicells.

Since the umuC locus actually consists of two genes, both of which are required for mutagenesis, an additional gene name must be introduced. We have consulted with T. Kato and suggest that the gene coding for the 16 kilodalton protein be termed umuD and the gene coding for the 45 kilodalton protein be termed umuC. Thus the umuC121::Mud(Ap, lac) mutation described previously (Bagg *et al.*, 1981) and the umuC122::Tn5 mutation discussed in this paper are correctly named. The umuC44 (Fig. 2-8) and umuC77 (data not shown) described by Kato and Shinoura (1977) are actually umuD44 and umuD77 alleles respectively. The exact nature of the umuC36 allele remains to be determined although it can be concluded that it is at least umuC<sup>-</sup>.

#### Complementation properties of the umuC/D locus on a high copy number plasmid

Having demonstrated the fact that the chromosomal alleles of umuC, including umuC36, were recessive, we were curious as to why our initial cloning efforts using pBR322 were unsuccessful. To investigate this problem we constructed a pBR322 derivative which carried the umuC<sup>+</sup> umuD<sup>+</sup> genes. Since we had no idea whether the genes could exist on a high copy number plasmid and, if so, whether they could then complement, we subcloned the genes by

linking them to a drug resistance marker and then subcloning the entire fragment on to a pBR322 derivative by selecting for the drug marker. We ligated EcoR1-KpnI cleaved pSE115 with EcoR1-KpnI cleaved pL10, a pBR322 derivative which carries the gpt gene. This was transformed into TK610 selecting for  $Km^R$ , which selected for the fragment from pSE115 that carried umuC and umuD, and for  $Ap^R$  which selected for the fragment of pL10 which carried the pBR322 origin of replication but not the gpt gene. Plasmid DNAs from the above selection were isolated and a derivative which contained the proper fragments was designated pSE117 (Fig 2-9). The frequency of occurrence of pSE117 molecules in this experiment was not unusual indicating that we had not selected for a variant of the umuC and umuD genes during the procedure. The pSE117 plasmid was then tested for the ability to complement strain TK610 using the plate screening method employed in our initial cloning attempts. Controls such as TK603 and TK610 (pSE115) appeared positive in this screen while TK610 and TK610 (pSE117) failed to yield any revertants on the plate patch screen.

To analyze the properties of pSE117 in greater detail, mutagenesis and survival curves were obtained for a variety of strains containing pSE117 (Figure 2-10). It is clear that at very low, uninducing doses of UV the non-mutability of the umuC36 allele is complemented. In fact, TK610(pSE117) and TK603(pSE117) are far more mutable than TK603 itself at these doses (0-0.5J), showing a gene dosage of umuC<sup>+</sup> effect for mutagenesis. However,

Figure 2-9. Schematic illustration of the construction of pSE117.

Figure 2-9

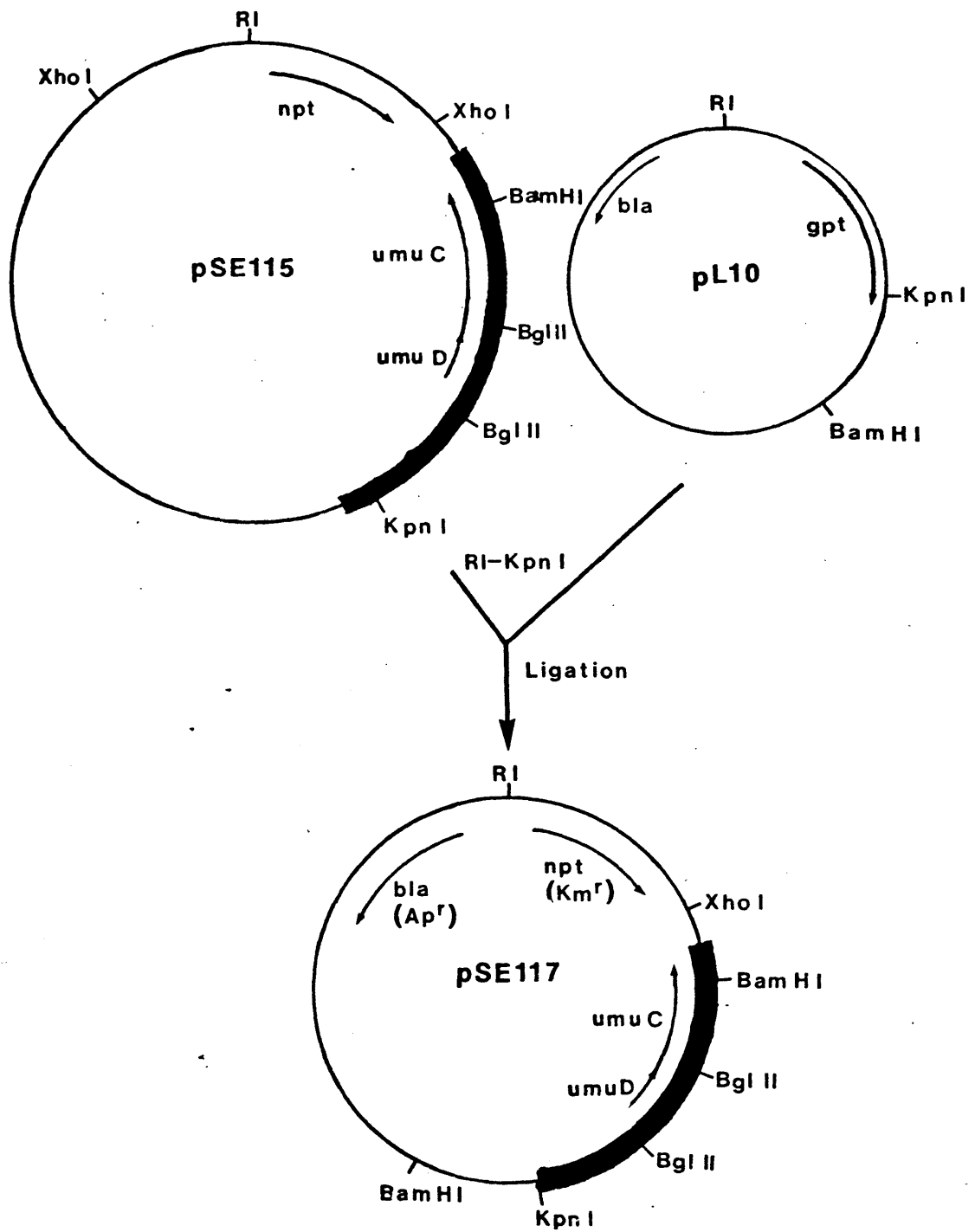
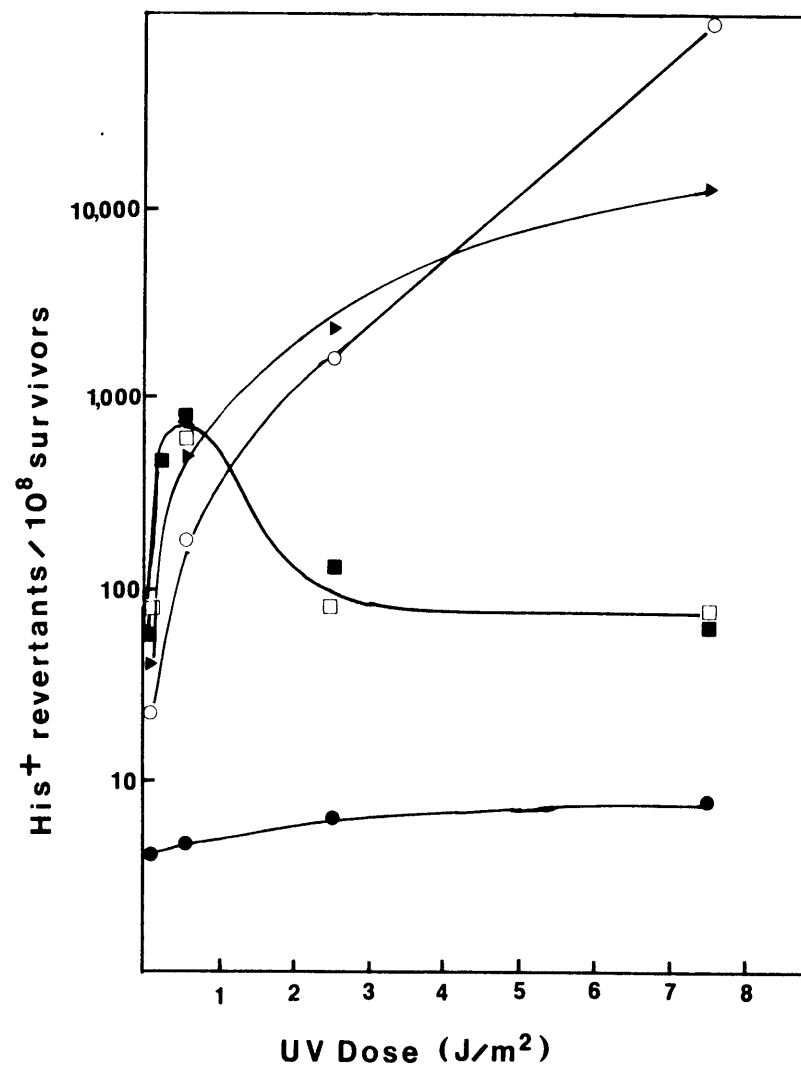
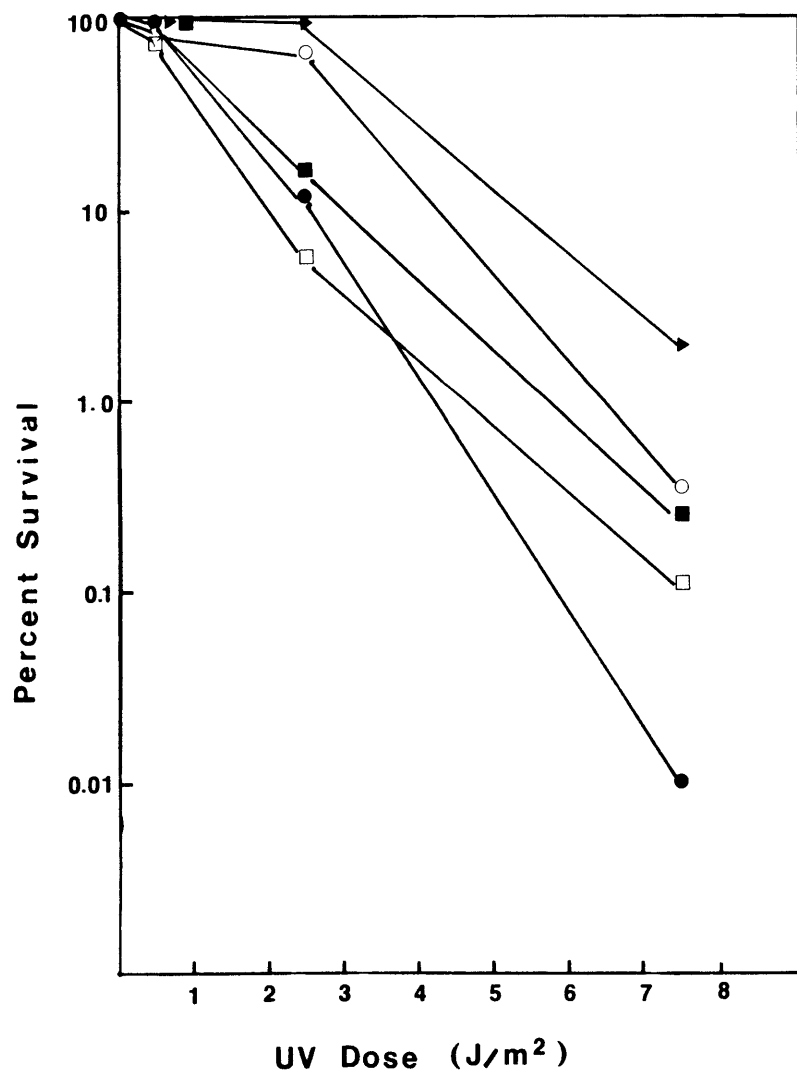


Figure 2-10. The effect of pSE117, a high copy number plasmid containing the umuC/D locus, on UV mutagenesis and survival in TK610 (uvrA<sup>-</sup>umuC<sup>-</sup>) and TK603 (uvrA<sup>-</sup>umuC<sup>+</sup>).

Cells were irradiated at a fluence of 0.1 J/m<sup>2</sup>/sec for various times and plated on M-9 minimal plates containing limiting histidine to measure both survival and mutagenesis as described in Materials and Methods. TK603, (○); TK610, (●); TK610(pSE115), (▶); TK603(pSE117), (□); TK610(pSE117), (■).



at doses large enough to induce the SOS functions we observed virtually no complementation of the TK610 strain. Furthermore, TK603(pSE117) looked nearly identical to TK610(pSE117) showing that the phenotype conferred upon the host bearing pSE117 was dominant. pSE117 also acted to decrease the UV-resistance of strains harboring it in a dominant fashion. These effects of pSE117 are in contrast to its low copy number parent plasmid pSE115, which increased mutability and UV-resistance to levels even greater than those displayed by wild type strains.

The presence of pSE117 appears to confer a UmuC-like phenotype on its host at doses of UV which cause induction of the SOS repair system. Thus it appears that overproduction of the umuC and umuD gene products is refractory to the proper functioning of the SOS repair system. This explains our inability to isolate a clone on pBR322 by screening for complementation of the umuC36 allele in TK610. At a UV dose of  $1.5-2 \text{ J/m}^2$ , mutability is decreased by a factor of 10 as is survival. This means that the number of revertants per patch would have been decreased by a factor of 100. Since wild type patches yield only 25 to 50 revertants, these clones would not appear positive. A low non-inducing dose of  $0.15-0.50 \text{ J/m}^2$  would have given us a positive signal in hindsight.

We have been unable to introduce pSE117 into DM1187, a recA441 (tif<sup>-</sup>) lexA51(Def) strain by transformation while control plasmids such as pBR322 and pL10 enter easily. We have been able



to introduce pSE117 into RB901, a  $\Delta$  (recA-srl)21lexA51(Def) strain. These facts point to a role for the recA protein in preventing the successful transformation of pSE117 into DM1187.

## Discussion

Having been unable to isolate the umuC locus by conventional cloning methods, we developed a cloning strategy which essentially gives one a positive selection for a gene for which there is normally no selection. By obtaining a Tn5 insertion in the umuC locus, we were able to isolate DNA homologous to that locus by cloning the kanamycin resistance gene from the umuC:Tn5 insertion strain. Using the chromosomal DNA adjacent to the Tn5 insertion, we probed a Charon 28 library of E. coli DNA and identified clones homologous to our probe. Using transient merodiploid complementation, we showed that one positive clone,  $\lambda_{SE14}$ , was able to complement the UmuC phenotype.

We subcloned the umuC locus from  $\lambda_{SE14}$  into a low copy number plasmid pSE101. By a combination of subcloning and Tn1000 insertion mutagenesis, we have identified a region of approximately 2.2kb that appears to contain the information necessary for the complementation of umuC<sup>-</sup> mutations.

We have shown that this region of DNA codes for two proteins with molecular weights of approximately 16,000 and 45,000 and that both of these proteins are required for UV mutagenesis. The genes for these two proteins appear to be organized in an operon that is repressed by the lexA protein. This constitutes the first formal identification of proteins that are specifically required for the cellular processes involved in UV and chemical mutagenesis in E. coli.

Overproduction of the umuC and umuD gene products appears to have a deleterious effect on the cells' ability to recover from UV damage. Cells bearing pSE117 present a umuC<sup>-</sup>-like phenotype when treated with UV doses large enough to induce SOS.

Several models could explain this effect. First, overproduction of these proteins could destroy a delicate stoichiometry required for error-prone repair, thus disabling the entire pathway and yielding a UmuC-like phenotype. A second model is that overproduction of these proteins could give excessive mutagenesis resulting in the frequent induction of lethal mutations in induced cells. A third model is that overproduction of these proteins is lethal for the induced cell, not because of too much mutagenesis, but because overproduction inhibits some aspect of proper cellular growth.

The first model is sufficient to explain the UV mutagenesis and survival data, however, it fails to explain the fact that strains that carry a null allele of lexA can maintain pSE117 only if they are deficient at the recA locus. The elimination of this functional pathway alone should not cause cell death because mutants deficient in this pathway have been isolated. All of this suggests that more is involved here than simply disabling the error-prone repair pathway.

The second model is more difficult to assess. It seems unlikely that I would fail to observe excessive mutagenesis in an analysis of mutants per survivors given the Poisson nature of the distribution of UV damage on the chromosome and the assumption that the mutagenesis

observed here is targeted. Alternatively, if when a cell bearing pSE117 is induced for SOS it induces massive, untargeted mutagenesis great enough to produce a lethal mutation in each induced cell, then the survivors would be viewed as uninduced cells. Still, among the survivors one would expect to observe mutagenesis levels at least as high as those observed at low, uninducing UV doses such as 0.15J and 0.5J. This is clearly not the case as mutagenesis levels decrease with increasing dose. Although induction of massive untargeted mutagenesis could explain the inability of a lexA(Def) strain to harbour pSE117, the evidence gathered here does not support this model.

I favor the third model. The fact that constitutive over-production of the umuC/D proteins in the presence of the recA<sup>+</sup> protein is lethal points toward an interference in a fundamental aspect of cell growth. Since it is likely that these proteins interact with the replication machinery, perhaps the replication apparatus is the target of interference. In order to explain the mutation frequency decline at inducing doses of UV, we must partially include the first model and suggest that while the umuC/D pathway is working to undermine cell growth, it is also no longer performing its normal function in induced cells.

Investigations of the properties of the umuC and umuD gene products should aid the elucidation of the mechanism of "error-prone repair" (Radman, 1975; Witkin, 1976). The most popular model to explain the molecular basis of mutagenesis has been that an

error-prone polymerase polymerizes past lesions that are normally refractory to DNA replication and introduced incorrect nucleotides (Boiteux et al., 1978; Lackey et al., 1982). In such a model the function of the umuC and umuD gene products could be to either code for components of a new polymerase or to code for proteins that modify an existing polymerase to make it error-prone. However, at this stage, other possible models cannot be eliminated and so must be considered as well. For example, mutagenesis could result from the action of an insertase that places bases into a apurinic or apyrimidinic sites. Such a model would be consistent with the conclusions of Kunkel et al. (1981) that apurinic sites can serve as premutagenic lesions. Another possibility would be that a "graftase" could ligate a short oligonucleotide patch over a lesion and thus allow polymerization to occur beyond the lesion. If the bases in the patch opposite the lesion were incorrect this would lead to a mutation. Yet another formal possibility would be that a single-stranded region of DNA containing a lesion could transiently pair with some other stretch of single-stranded DNA via limited homology on either side of the lesion. Removal of the lesion via mismatch repair followed by separation of the strands could result in an incorrect base being introduced at the site of the lesion. It is certain that the umuC and umuD proteins do not act alone to produce mutagenesis. A functional recA protein is also needed for error-prone repair (Little and Mount 1982, J. Kruger, unpublished results, see intro to this thesis). The fact that no other umu loci have been found to date suggests that the secondary role played by recA is not to induce the expression of another gene necessary for error-prone

repair, but actually to participate structurally in the repair process. The lethality of overproduction of the umuC and umuD proteins in the presence of the recA protein suggest that they work together in some capacity.

I would like to propose a model for the role of recA and umuC and umuD proteins in error-prone repair that attempts to tie together a number of observations by scientists in the field. When the DNA polymerase encounters a lesion such as a pyrimidine dimer, the replication fork stalls while the SOS repair system is induced. Due to lagging strand synthesis and DNA unwinding, ahead of the stalled replication fork is single-stranded gaps of DNA. It is my hypothesis that when the recA protein is induced it binds these single-stranded gapped regions at the replication fork. When bound to single-stranded DNA, the recA protein is conformationally altered such that its protease function is active. In this conformation the umuC and umuD proteins, positioned at the replication fork, can interact with the polymerase to inhibit proofreading or alter base selection or whatever to ensure polymerization beyond the dimer. This model has several advantages: 1) it explains the secondary role of the recA protein in mutagenesis, 2) it directs the umuC and umuD proteins to the sites of DNA damage where they can interact with the polymerase and 3) it predicts targeted mutagenesis. The main prediction generated by this model is that umuC or umuD proteins, or both, interact with the recA protein when it is bound to single-stranded DNA. Having cloned the umuC/D genes will allow us to test this prediction using radiolabeled proteins.

Other speculative models could be discussed as well but it is our expectation that the identification of the umuC and umuD gene products described in this paper will allow us to carry out a number of systematic biochemical studies of the cellular processes involved in UV and chemical mutagenesis in E. coli and thus determine the actual mechanism.

## References

- Backman, K. and Ptashne, M. (1978). Cell **13**, 65-71.
- Bagg, A., Kenyon, C.J. and Walker, G.C. (1981). Proc. Natl. Acad. Sci. USA **78**, 5749-5753.
- Benton, W.D. and Davis, R.W. (1977). Science **196**, 180-182.
- Boiteux, S., Villani, G., Spadari, S., Zambrano, F. and Radman, M. (1978). In DNA Repair Mechanisms (Hanawalt, P.C., Friedberg, E.C. and Fox, C.F., eds). pp. 73-84. Academic Press, New York.
- Bonner, W.M. and Laskey, R.A. (1974). Eur. J. Biochem. **46**, 83-88.
- Brent, R. and Ptashne, M. (1980). Proc. Nat. Acad. Sci. USA **77**, 1932-1936.
- Brent, R. and Ptashne, M. (1981). Proc. Natl. Acad. Sci. USA **78**, 4204-4208.
- Broker, T.R. (1977). In DNA Insertion Elements, Plasmids, and Episomes (Bukliari, A.I., Shapiro, J.A. and Adhya, S.L., eds.) pp. 403-408. Cold Spring Harbor Laboratory, New York.
- Casadaban, M.J., Chou, J. and Cohen, S.N. (1979). J. Bacteriol. **143**, 971-980.
- Chung, S., Landfear, S.M., Blumberg, D.D., Cohen, N.S. and Lodish, H.F. (1981). Cell **24**, 785-797.
- Cohen, S.N. and Chang, A.C.Y. (1977). J. Bacteriol. **132**, 734-737.
- Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972). Proc. Nat. Acad. Sci. U.S.A. **69**, 2110-2114.
- Fogliano, M. and Schendel, P.F. (1981). Nature (London), **289**, 196-198.
- Holmes, D.S. and Quigley, M. (1981). Anal. Biochem. **114**, 193-197.
- Huisman, O. and d'Ari R. (1981). Nature (London), **290**, 797-799.
- Jorgensen, R., Rothstein, S.J. and Reznikoff, W.S. (1979). Molec. Gen. Genet. **177**, 65-72.
- Kato, T. and Shinoura, Y. (1977). Gen. Genet. **156**, 121-131.
- Kenyon, C. and Walker, G. (1980). Proc. Nat. Acad. Sci. USA **77**, 2819-2823.
- Kenyon, C. and Walker, G. (1981). Nature (London), **289**, 808-810.
- Kunkel, J.A., Shearman, C.W. and Loeb, L.A. (1981). Nature (London), **291**, 349-351.
- Lackey, D., Kruass, S.W. and Linn, S. (1982). Proc. Natl. Acad. Sci. USA **79**, 330-334.
- Little, J.W. and Mount, D.W. (1982).
- Little, J.W., Mount, D.W. and Yanisch-Perron, C.R. (1981). Proc. Natl. Acad. Sci. USA **78**, 4199-4203.
- Miller, H.I., Kirk, M. and Echols, H. (1981). Proc. Nat. Acad. Sci. USA **78**, 6754-6758.
- Mount, D.W., Low, K.B. and Edmiston, S. (1972). J. Bacteriol. **112**, 886-893.
- Mount, D.W. (1977). Proc. Nat. Acad. Sci. USA **74**, 300-304.
- Perry, K.L. and Walker, G.C. Nature, **300**, 278-281.
- Radman, M. (1975). In Molecular Mechanisms for Repair of DNA (Hanawalt, P. and Stelow, R.B., eds.) pp. 355-367. Plenum Press, New York.
- Rigby, P.W.J., Deckman, M., Rhodes, C. and Berg, P. (1977). J. Mol. Biol. **113**, 237-254.
- Sancar, A. and Rupp, W.D. (1979). Biochem. Biophys. Res. Commun. **90**, 123-129.



- Sancar, A., Wharton, R.P., Seltzer, S., Kacinski, B.M., Clarke, N.D., and Rupp, W.D. (1981). J. Mol. Biol. 148, 45-62.
- Sancar, G.B., Sancar, A., Little, J.W. and Rupp, W.D. (1982). Cell, 28, 523-530.
- Sharp, P.A., Sugden, B. and Sambrook, J. (1973). Biochemistry 12, 3055-3063.
- Shih, C. and Weinberg, R.A. (1982). Cell 29, 161-169.
- Sninsky, J.J., Uhlin, B.E., Gustafsson, P. and Cohen, S.N. (1981). Gene 16, 275-286.
- Southern, E.M. (1975). J. Mol. Biol. 98, 503-517.
- Steinborn, G. (1978). Mol. Genet. 165, 87-93.
- Taylor, D. and Cohen, S. (1979). J. Bacteriol. 137, 92-104.
- Timmis, K., Cabello, F. and Cohen, S.N. (1978). Mol. Gen. Genet. 162, 121-137.
- Walker, G.C. and Dobson, P.P. (1979). Molec. Gen. Genet. 172, 17-24.
- Witkin, E.M. (1976). Bacteriol. Rev. 40, 869-907.

## Chapter III

### Introduction

pKM101 is a 34.5 kilobase (kb) N-incompatibility plasmid that was derived from the clinically-isolated plasmid R46 by a series of in vivo manipulations (Mortelmans and Stocker, 1979; Langer and Walker, 1981). It is one of a number of plasmids that increase the susceptibility of Escherichia coli and Salmonella typhimurium to mutagenesis by a variety of agents including UV and also increase the resistance of these bacteria to killing by UV (Chernin and Mikoyan, 1981). The capacity of pKM101 to enhance mutagenesis is closely associated with its ability to enhance recovery from UV damage since single mutations have been isolated which affect both processes (Walker, 1978; Shanabruch and Walker, 1980). The ability of pKM101 to increase chemical mutagenesis led to its introduction into the Ames Salmonella tester strains for the detection of carcinogens as mutagens and it has played a major role in the success of the system (McCann et al., 1975).

Mutations at four different chromosomal loci - recA, lexA, umuC, and umuD - can make E. coli nonmutable by UV and a variety of chemical agents (Miura and Tomezawa, 1968; Mount et al., 1972; Kato and Shinoura, 1977; Elledge and Walker, 1983). The recA<sup>+</sup> and lexA<sup>+</sup> gene products are involved in the regulation of the set of SOS genes in E. coli that are induced by DNA damage (Witkin, 1976; Mount et al., 1972, Kenyon and Walker, 1980; Little and Mount, 1982) so that mutations at these loci are very pleiotropic. In contrast, umuC and umuD mutants are nonmutable with many agents and are deficient in Weigle-reactiva-

tion but still exhibit the rest of the SOS responses (Kato and Shinoura, 1977; Steinborn, 1978; Elledge and Walker, 1983). Thus umuC and umuD are the best candidates for genes whose products are uniquely involved in "error-prone repair" (Kato and Shinoura, 1977; Elledge and Walker, 1983).

pKM101 and R46 are able to suppress the nonmutability of umuC/D mutants (Walker and Dobson, 1978; Steinborn, 1978), observations which suggested the possibility that these plasmids carried analogs of the umuC/D genes (Walker and Dobson, 1979; Shanabruch and Walker, 1980). This has recently been confirmed by our demonstration that the muc region of pKM101 coded for two proteins, mucA and mucB (Perry and Walker, 1982) whose molecular weights and transcriptional organization were similar to those of the chromosomally-encoded umuD and umuC gene products respectively (Elledge and Walker, 1983).

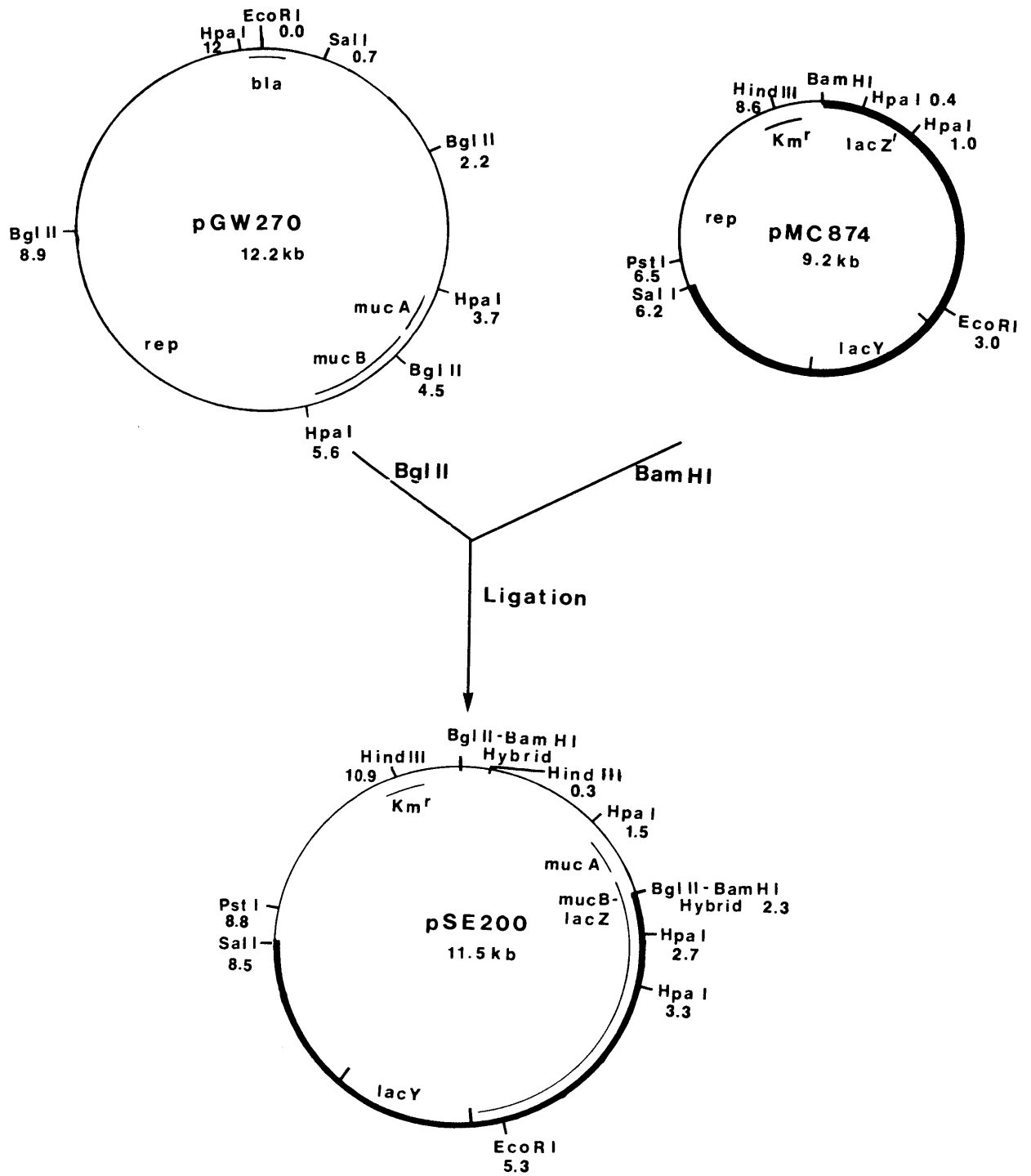
The fact that pKM101 and R46 are unable to enhance mutagenesis or increase resistance to UV killing in a recA or lexA ( $\text{Ind}^-$ ) background (Walker, 1977) suggested that the muc genes of pKM101 might be part of the recA<sup>+</sup>lexA<sup>+</sup> regulatory network. Furthermore, certain observations such as the enhancement of Weigle-reactivation by pKM101 (Walker, 1978) and the synergistic effects of pKM101 and tif mutations (Walker, 1977; Doubleday et al., 1977) had indicated that some of the pKM101-mediated effects were inducible. Therefore we decided to use a lac fusion approach to directly investigate whether the muc genes of pKM101 were induced by DNA damage.

## RESULTS

In order to facilitate our investigation of the regulation of the muc genes of pKM101 we wished to place lacZ, the structural gene for  $\beta$ -galactosidase, under the transcriptional control of the muc promoter and regulatory region. Since protein fusions offer some additional experimental advantages over operon fusions, we decided to construct a fusion which would result in the production of a hybrid protein consisting of an enzymatically active segment of  $\beta$ -galactosidase joined to the amino terminus of a muc gene product. In attempting the construction of such a fusion we decided to try to take advantage of a BglIII site located in the mucB region (Fig. 3-1) which could form a hybrid site with the BamHI site of pMC874, a protein fusion vector engineered to have a BamHI site early in the lacZ gene (Casadaban and Cohen, 1980). If the mucB protein was in the correct reading frame at the BglIII site we would be able to obtain the desired protein fusion.

The construction of the fusion is outlined in Fig. 3-1. pGW270, a deletion derivative of pKM101::Tn5 $\Omega$ 735 (Langer *et al.*, 1981) which retains all of the mutagenesis and UV resistance properties of pKM101 was cleaved with BglIII and ligated with BamHI-cleaved pMC874. After transformation into a GW1000 strain kanamycin-resistant ( $Km^R$ ) colonies were selected on medium containing Xgal and ones that were blue after one day were chosen for further study. Since the lacZ gene of pMC874 lacks the DNA coding for the first seven amino acids of  $\beta$ -galactosidase the blue colonies had to contain pMC874 derivatives whose inserts supplied not only a promoter but also translational start signals as

Fig. 3-1. Schematic illustration of the construction of pSE200. The heavy lines are sequences of the lacZ and lacY genes of pMC874. The map of pMC874 is redrawn from Casadaban and Cohen (1980).



well as DNA coding for the amino terminus of a protein that is in the same reading frame as  $\beta$ -galactosidase.

These strains were then examined for their ability to induce synthesis of  $\beta$ -galactosidase after treatment with  $15 \text{ J/m}^2$  of UV irradiation. Several of these did induce and their plasmid DNAs were subjected to EcoRI-PstI and HindIII restriction endonuclease cleavage analysis. All derivatives tested gave identical restriction fragment patterns showing insertion of the 2.3 kb BglII fragment of pGW270 into the BamHI site of pMC874. Furthermore, the orientation of the inserted fragment was such that the BglII site interior to the muc region was ligated to BamHI site in the lacZ gene as illustrated in Fig. 3-1. One of the derivatives tested was chosen for further analysis and was designated pSE200.

The fact that the mucB-lacZ fusion we constructed expressed  $\beta$ -galactosidase activity indicated that the mucB protein was in the same reading frame as  $\beta$ -galactosidase at the BglII-BamHI junction and therefore established the reading frame of the mucB protein. Furthermore this fusion indicated that the direction of transcription of the mucB gene is clockwise on the pGW270 map and hence on clockwise our pKM101 map (Langer *et. al.*, 1981). Finally, these experiments demonstrated that the mucB gene product is induced by treatment of cells with a DNA damaging agent.

### Induction of muc expression by other agents

Because the induction of the SOS functions can be triggered by a diverse array of DNA damaging agents (Witkin, 1976), we were interested in the effects of other SOS inducers on the synthesis of  $\beta$ -galactosidase in strains carrying pSE200. As illustrated in Table 3-1, methyl methanesulfonate (MMS), 4-nitroquinoline-1-oxide (NQO), N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), and mitomycin C all proved to be effective inducers of the mucB gene product.

### Effects of recA, lexA(Ind<sup>-</sup>) and lexA(Def) mutations on mucB expression

The mutability of cells containing pKM101 is dependent upon the function of the recA and lexA proteins (Walker, 1977). These proteins are known to regulate the expression of a variety of damage-inducible genes (Kenyon and Walker, 1980; Little and Mount, 1982). The lexA protein has been demonstrated to directly repress transcription of the promoters of several cloned damage-inducible in vitro (Little and Mount, 1982; Kenyon et al., 1982). Induction of these genes occurs when recA protein proteolytically cleaves the lexA protein in response to DNA damage (Little and Mount, 1982). Since the mucB gene was damage-inducible, we wanted to determine whether it was regulated by the recA and lexA proteins on a similar fashion. We therefore introduced pSE200 into strains bearing either the recA56 mutation or the lexA3 (Ind<sup>-</sup>) mutation. As shown in Fig. 3-2, UV induction of  $\beta$ -galactosidase activity proceeds normally in an AB1157 (recA<sup>+</sup>lexA<sup>+</sup>) background. However the presence of either the recA56 allele or the



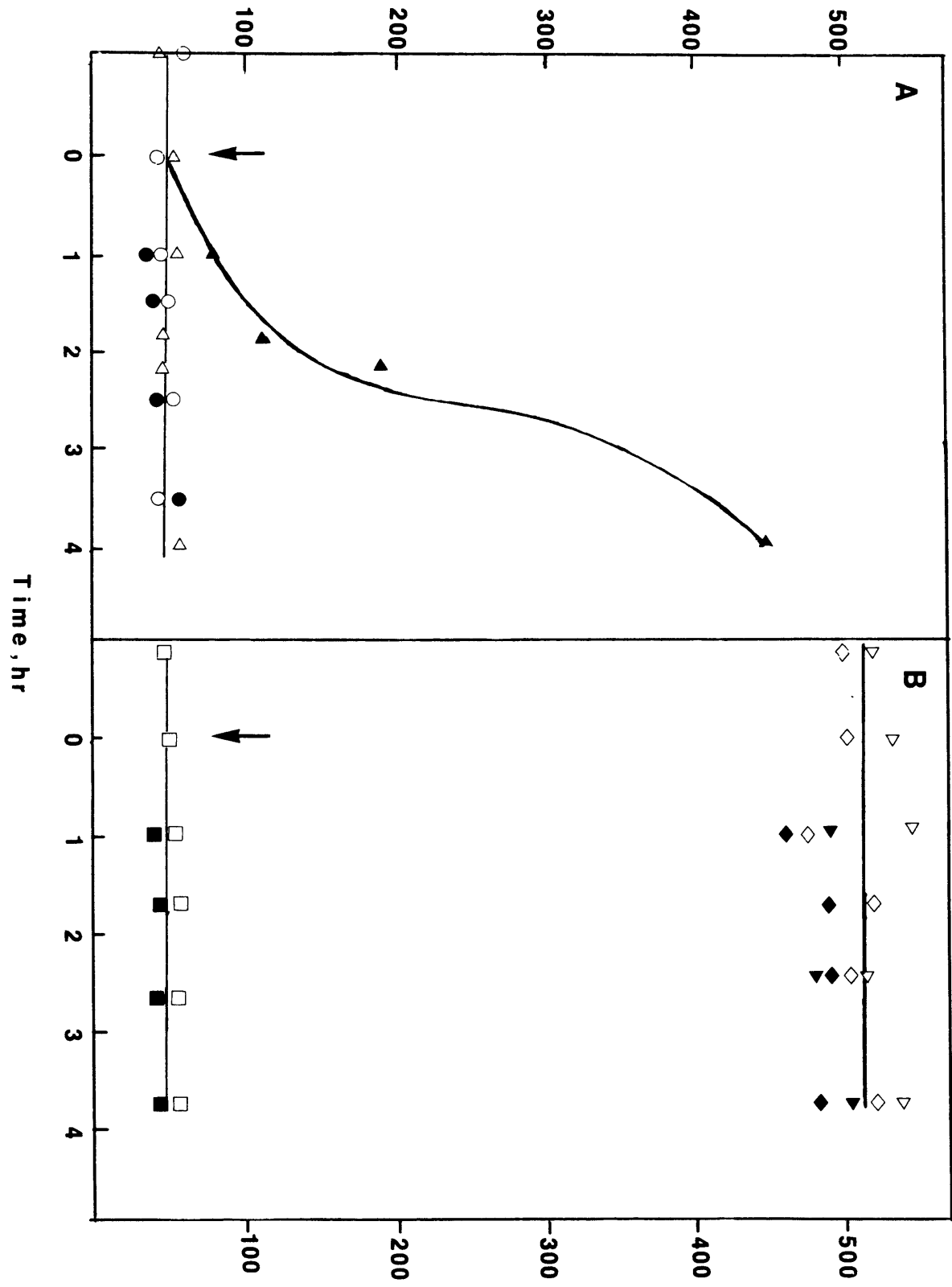
Table 3-1. Induction of  $\beta$ -galactosidase in a mucB-lacZ fusion strain by various agents.

Chemical	Concentration	Units of $\beta$ -galactosidase activity/OD <sub>600</sub>
-	-	70
mitomycin C	0.25 $\mu$ g/ml	142
4-nitroquinoline-1-oxide	15 $\mu$ g/ml	503
methyl methanesulfonate	.03 $\mu$ g/ml	194
N-methyl-N'-nitro-N-nitrosoguanidine	1 $\mu$ g/ml	250

<sup>a</sup> An exponentially growing culture of AB1157 (pSE200) in minimal media was split into several aliquots. One aliquot was untreated, and chemicals were added to others to the concentration shown. The cells were then incubated at 37<sup>o</sup> for 4 hrs and the  $\beta$ -galactosidase activity was determined.

Fig. 3-2. Kinetics of UV induction of  $\beta$ -galactosidase in various genetic backgrounds containing pSE200. Cells were grown in supplemented minimal media at 37<sup>o</sup>. Cells were treated with 15 J/m<sup>2</sup> of UV irradiation at the times indicated by the arrow. Aliquots (1 ml) were removed periodically and the total activity of  $\beta$ -galactosidase present in the cultures was determined. A. AB1157(pSE200) without UV ( $\Delta$ ), with UV ( $\blacktriangle$ ); DM49(pSE200)(lexA3 Ind<sup>-</sup>) without UV (O), with UV ( $\bullet$ ). B. JC2926(pSE200)(recA<sup>-</sup>) without UV ( $\square$ ), with UV ( $\blacksquare$ ); DM1180(pSE200)(lexA(Def)), without UV ( $\triangleright$ ), with UV  $\blacktriangleright$ ; DM1415(pSE200)(recA<sup>-</sup> lexA(Def)) without UV ( $\diamond$ ), with UV ( $\blacklozenge$ ).

$\beta$ -Galactosidase activity  
units/OD<sub>600</sub> unit



lexA3 (Ind<sup>-</sup>) allele totally abolishes the induction of  $\beta$ -galactosidase activity by UV light.

To further analyze this recA<sup>+</sup>lexA<sup>+</sup>-dependence, we examined the expression of mucB in a strain bearing a null allele of lexA termed lexA(Def) (formerly called spr). The presence of the lexA(Def) allele rendered  $\beta$ -galactosidase synthesis constitutive at high levels and no further increase was observed after UV irradiation. Thus the lexA protein appears to regulate the mucB gene product in a negative fashion. Furthermore, when we introduced pSE200 into a strain that was recA lexA(Def), we again observed high-level constitutive synthesis of  $\beta$ -galactosidase that was not increased after irradiation. This indicates that, in the absence of the lexA protein, recA protein is not needed for expression of the mucB gene product. These observations suggest that the lexA protein acts as the direct repressor of the mucB gene product and thus that the recA protein acts to lower levels of lexA protein by proteolytically cleaving the lexA protein in response to DNA damage.

#### Effect of the tif mutation on mucB expression

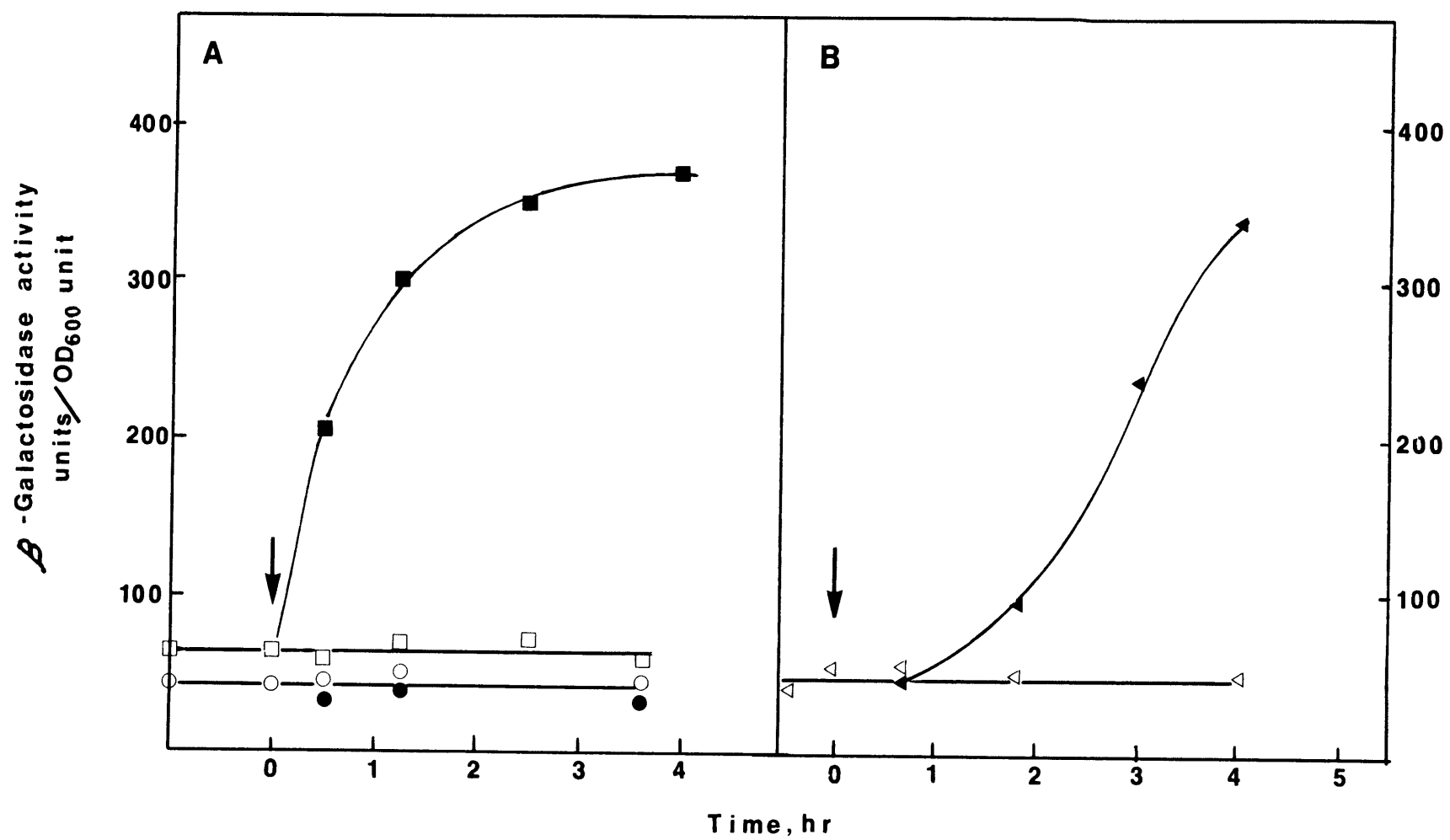
If the mucB gene is indeed controlled as outlined above then we expected that it would be induced in the absence of DNA damage in a tif background by shifting the culture from 30°C to 42°C and adding adenine, a treatment known to induce SOS functions. As anticipated, the tif allele allowed induction of  $\beta$ -galactosidase at 42°C in the presence of adenine and this induction was blocked in the tif lexA3 double mutant (Fig. 3-3).

Fig. 3-3. Analysis of the kinetics of induction of  $\beta$ -galactosidase activity in strains bearing mutations at the recA locus.

A. GC3217(pSE200) [recA441(tif-1)] at 30°C (□), at 42°C plus adenine (■); DM1180(pSE200) (recA441 (tif-1) lexA3 (Ind<sup>-</sup>)) at 30°C (○), at 42°C plus adenine (●).

B. JC10521 (recA430 (lexB30)) (pSE200) without UV (Δ), with UV (▲).

The arrow in A. indicates when the cultures were shifted to 42° C and adenine was added to a concentration of 100  $\mu$ g/ml. The arrow in B. indicates the time at which the cells were treated with 15 J/m<sup>2</sup> of UV irradiation.



### Effect of the *recA430*(*lexB30*) mutation on *mucB* expression

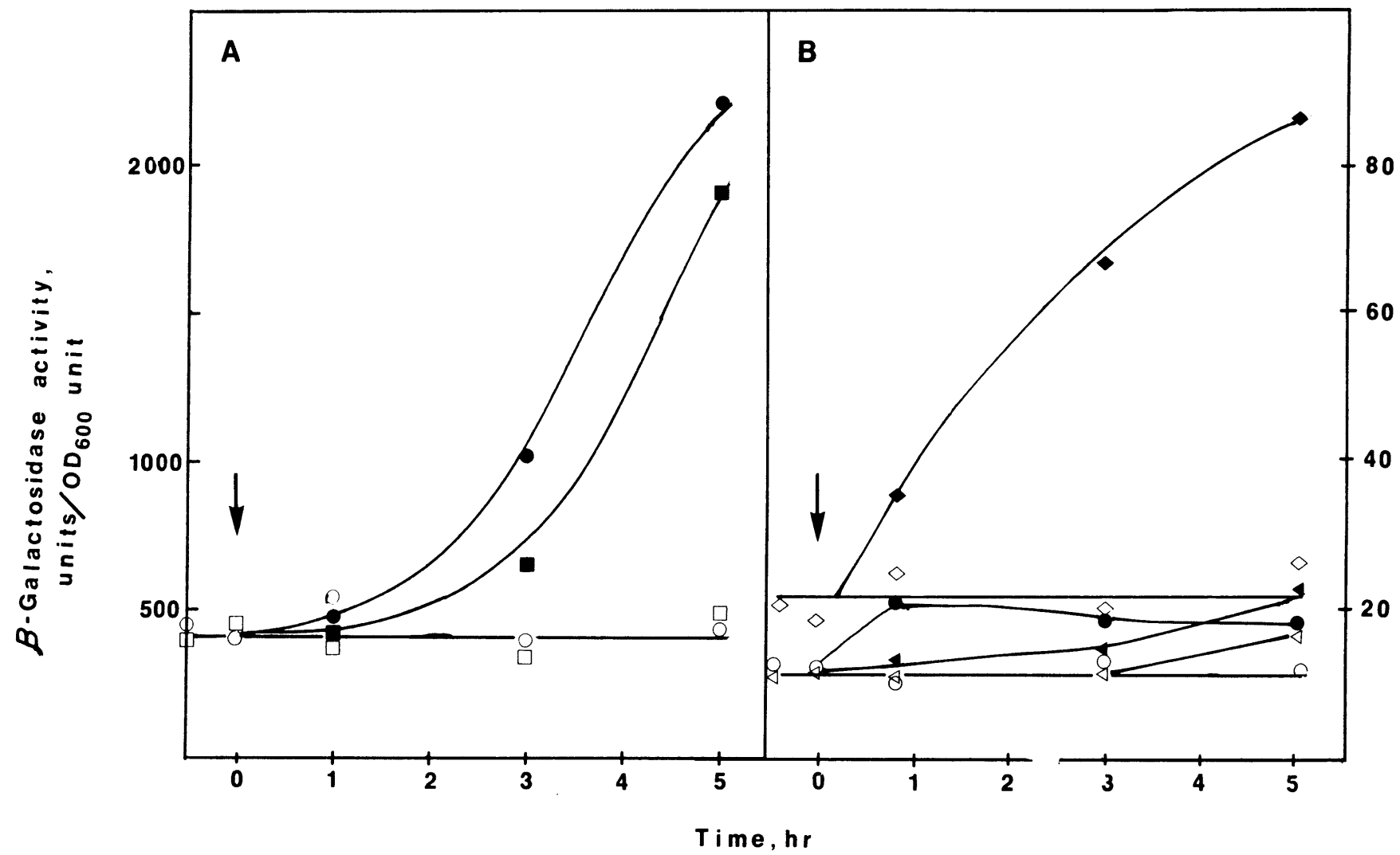
Another mutation affecting the regulation of SOS phenomena is *recA430*, formerly called *lexB30*. This allele of *recA* was found to be deficient in induction of prophage lambda, deficient in UV and chemical mutagenesis, UV sensitive, but recombination proficient (Devoret and Blanco, 1970, Morand *et al.*, 1977). The *recA430* protein was found to lack the ability to cleave the lambda repressor *in vitro* (Roberts and Roberts, 1981). Thus it appeared as if *recA430* was a mutant deficient in the protease activity of the *recA* protein and its phenotype explained by its loss of the ability to cleave the *lexA* repressor. We introduced pSE200 into JC10521, a strain bearing the *recA430* mutation and found that  $\beta$ -galactosidase activity was induced normally in this *recA430* background (Fig. 4). We reasoned that either the *recA430* protein retained the ability to cleave the *lexA* protein or that induction was due to the presence of the *mucB-lacZ* fusion on the relatively high copy number plasmid vector. For example, the multiple copies of the *muc* operator sequence could have been titrating out *lexA* protein and aiding in the induction process or else perhaps in a *recA430* background UV irradiation could have led to an increase in the copy number of the plasmid.

To investigate these classes of possibilities we decided first to see if we could reproduce the phenomenon using another *din-lac* fusion. The *recA430* allele was introduced into GW1030 (Kenyon and Walker, 1980), a strain containing the *dinB::Mud*(Ap, *lac*) insertion to give

Fig. 3-4. Analysis of the differential effects of the recA430 mutation on the induction of the dinB::Mu dl(Ap<sup>R</sup>, lac) fusion located in the chromosome (GW1039) or on a multicopy plasmid (pGW510).

A. AB1157(pGW510)(recA<sup>+</sup>) without UV (□), with UV (■);  
 JC10521(pGW510) (recA430) without UV (○), with UV (●). B. GW1030  
 [dinB::Mu dl(Ap, lac)] without UV (◇), with UV (◆); GW1039  
 [dinB::Mu dl(Ap, lac) recA430] without UV (Δ), with UV (▲);  
 GW1039(pSE210) without UV (○), with UV (●).





strain GW1039; we also introduced pGW520 (Kenyon et al., 1982) a pBR322 derivative carrying the dinB-lacZ fusion of GW1030 into the recA430 strain JC10521. We then tested these strains for UV inducibility of  $\beta$ -galactosidase activity and found that it was inducible in JC10521(pGW520) but uninducible in GW1039. Thus, even though the dinB-lacZ fusion is normally not inducible in a recA430 strain it resembles the mucB-lacZ fusion in being inducible in a recA430 background when cloned onto a multicopy vector. We also ruled out a possible interference of the dinB mutation itself on the recA430 protein's ability to cleave the lexA protein by testing the inducibility of GW1039(pSE200);  $\beta$ -galactosidase activity was now induced by UV-irradiation in this strain to the same extent as in JC10521(pSE200) (data not shown).

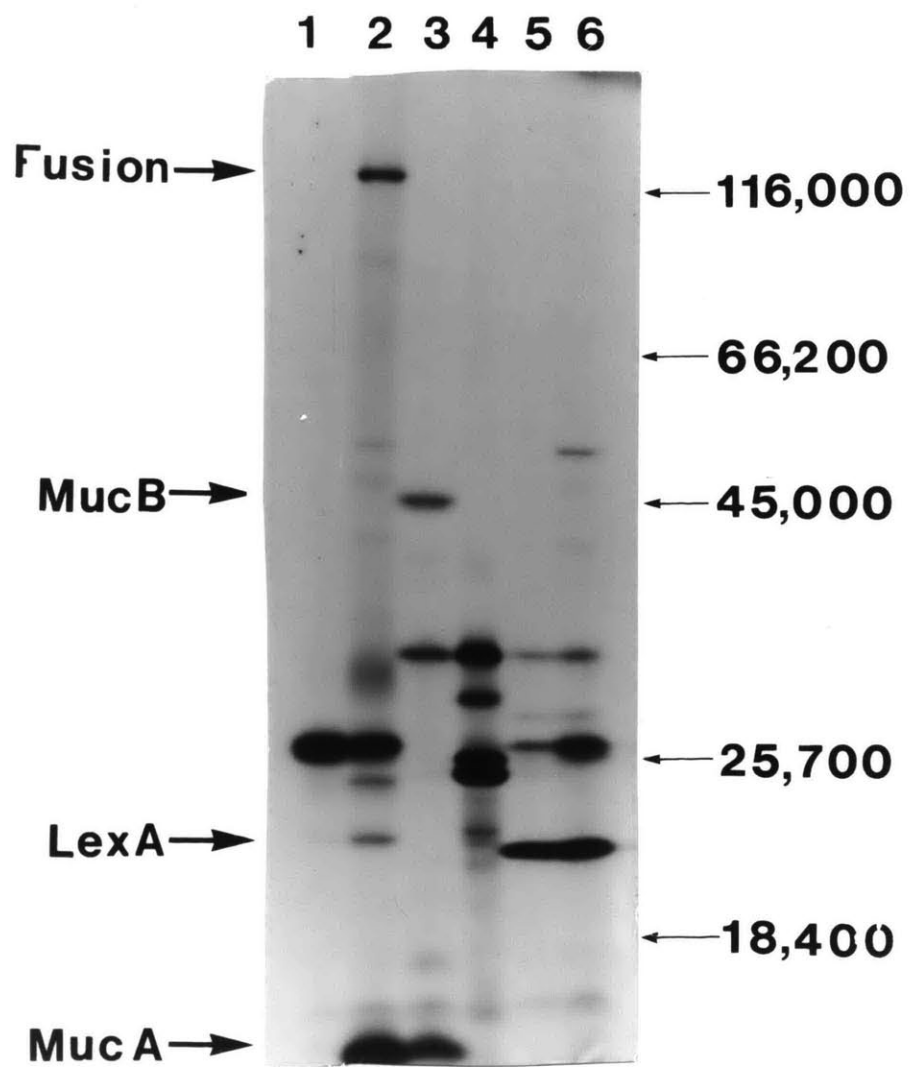
To test the hypothesis that multiple copies of the lexA-binding operator sequences were aiding the induction of  $\beta$ -galactosidase activity in strains containing pSE200 or pGW520, we constructed a chloramphenicol resistant derivative of pGW520 which retained the operator-promoter region (Kenyon et al., 1982), but which was deleted for the lacZ gene. This plasmid, pSE210, was introduced into GW1039 [dinB::Mud(Ap, lac) recA430] and assayed for UV-induced  $\beta$ -galactosidase activity. The strain remained uninducible for  $\beta$ -galactosidase activity therefore ruling out the possibility that multiple copies of the operator were titrating out most of the lexA protein and thereby facilitating the induction.

We were also able to rule out the possibility that the induction of  $\beta$ -galactosidase activity in recA430(pSE200) and recA430(pGW520) strains was due to an increase in plasmid copy number after UV. This was done by constructing an operon fusion of lacZ to the Tc promoter on pBR322 and showing that the  $\beta$ -galactosidase activity was not induced by UV in recA430 strains containing this plasmid (data not shown).

#### Analysis of regulation using maxicells

Our analysis of the regulation of the mucB gene suggested that the lexA protein was acting as the direct repressor of the mucB gene and we were interested in seeing whether independent evidence for such regulation could be obtained using maxicells. Thus we examined proteins synthesized from the plasmid pSE200 in maxicells prepared from either a recA lexA<sup>+</sup> strain or a recA lexA(Def) strain but could not observe a quantitative difference in muc protein levels relative to control proteins (data not shown). Since it seemed possible that the levels of lexA might be quite low in maxicells we introduced the plasmid pSE152, a spectinomycin-resistant derivative of pRB160 which carries the lexA<sup>+</sup> gene (Elledge and Walker, 1983) into pSE200-bearing strains. The protein products synthesized in maxicells derived from these strains are shown in Fig. 3-5. The pSE200 plasmid codes for the mucA protein in addition to the hybrid mucB- $\beta$ -galactosidase protein. In the presence of the lexA<sup>+</sup> plasmid pSE152 both the mucA and protein and hybrid mucB- $\beta$ -galactosidase protein were synthesized in very low levels relative to the strains lacking pSE152. Thus it appears that

Fig. 3-5. Repression of the synthesis of the mucB-lacZ protein fusion and the mucA gene product by the lexA protein in maxicells. Plasmid-encoded proteins were labeled by the maxicell method and were analyzed on 10-20% SDS polyacrylamide gels as described previously (Elledge and Walker, 1983). Lane 1, RB901(pMC874); lane 2, RB90(pSE200); lane 3, RB901(pGW1700); lane 4, RB901(pKB354); lane 5, RB901(pSE152); lane 6, RB901(pSE200)(pSE152).



the mucA and mucB genes are coordinately regulated in a manner consistent with the lexA protein being the direct repressor. Furthermore, Tn1000 insertions which eliminate the mucA protein also eliminate the mucB protein (Perry and Walker, 1982). Together with the coordinate regulation of mucA and mucB these results suggest that these two genes are organized in an operon with the mucA gene transcribed first.

## DISCUSSION

Observations such as the enhancement of Weigle-reactivation by pKM101 (Walker, 1978) and the synergistic effects of pKM101 and tif mutations (Walker, 1977; Doubleday et al., 1977) have indicated that at least some of the effects of pKM101 on mutagenesis and recovery from DNA damage are inducible. In the work described in this paper we have shown that this inducibility is due, at least in part, to the induction of the mucA and mucB gene products.

The genetic analyses we have carried out are consistent with the mucA and mucB genes being organized in an operon that is directly repressed by the lexA protein. Thus the lexA protein appears to function not only as the repressor of a large set of chromosomal genes - recA, lexA, uvrA, uvrB, uvrC, uvrD, umuC, umuD, sulA, hima, ruv, dinA, dinB, dinD, and dinF - but also as the repressor for genes found on naturally occurring plasmids - the mucA and mucB genes of pKM101 and the colicin gene of colE1.

We had hoped to investigate the possibility that pKM101 coded for a function or functions that were involved in the regulation of the muc locus by placing both pKM101 and pSE200 in a lexA(Def) strain. However we were unable to introduce pKM101 into a lexA(Def) strain by either mating or transformation. We do not yet know whether this was because pKM101 killed the lexA(Def) strain or if the plasmid was unable to be stably maintained in that background. We also attempted to construct the same mucB-lacZ fusion by cloning the 5.8 kb BglIII B

fragment from pKM101 (Langer and Walker, 1981) (BglIII B contains the mucB gene up to the BglIII site, the mucA gene, and additional pKM101 DNA from the counterclockwise side of the muc genes). Although we repeatedly obtained several protein fusions, none of these plasmids contained the BglIII B fragment alone. This would appear to indicate that, in the absence of other pKM101 functions, the BglIII B fragment of pKM101 cloned into pMC874 is lethal either for the cell or for the plasmid itself.

The parallel between the mucA/B locus and the umuD/C locus is striking. Both loci code for two polypeptides with molecular weights of 16 and 45 kilodaltons respectively (Perry and Walker, 1982; Elledge and Walker, 1983). Moreover, since the mucA/B genes can genetically suppress mutations in the umuD/C genes (Walker and Dobson, 1979) it seems likely that the two sets of proteins have similar functions. Both loci seem to be organized in an operon with the smaller protein transcribed first and, as we have shown here, the mucA/B genes appear to be repressed by the lexA protein as are the umuD/C genes (Bagg *et al.*, 1981; Elledge and Walker, 1983). This evidence strongly supports the conclusion that these two loci, one on the chromosome and one on a plasmid, have a common evolutionary origin. The two loci must have diverged considerably, at least at the nucleic acid sequence level, since they failed to cross-hybridize in Southern hybridization experiments (Elledge and Walker, unpublished results).

One of the most interesting implications of this study is that concerning the nature of the recA430(lexB30) allele (Devoret and



Blanco, 1970). This allele has been shown to render cells UV-sensitive, and deficient in mutagenesis, Weigle-reactivation, and the ability to induce prophage  $\lambda$  (Devoret and Blanco, 1970; Morand *et al.*, 1977; Roberts and Roberts, 1981). Purified recA430 protein has also been shown to be unable to cleave the lambda repressor *in vitro* (Roberts and Roberts, 1981) and thus the effect of the recA430 allele seemed to make the recA protein defective in its protease activity.

In this study, in the course of trying to understand the effects of recA430 on induction of the muc locus, we demonstrated that recA430 blocks the UV induction of a chromosomally-located dinB-lacZ fusion but fails to block induction of the identical fusion cloned onto pBR322. We have ruled out a number of possible explanations for this phenomenon including: i) titration of lexA protein due to multiple copies of a lexA binding sequence on the high copy number plasmid ii) an increase in copy number of pBR322 after UV and iii) a requirement for dinB<sup>+</sup> function to induce din genes in a recA430 background.

Our observations are therefore consistent with a model in which the recA430 protein is able to cleave the lexA protein but does not cleave it as efficiently as the recA<sup>+</sup> protein. Thus, in a recA430 strain, the intracellular lexA concentrations would only be reduced to intermediate levels by the same inducing dose that would lead to very low lexA levels in a recA<sup>+</sup> background. At such intermediate lexA concentrations, sets of lexA-repressed genes with  $K_d$ 's for lexA protein dissociation from their operators greater than the repressor concentration would have been induced, while sets of lexA-repressed

genes with  $K_d$ 's less than the lexA concentration would remain uninduced. An explanation for the observed difference in expression of the chromosomal dinB-lacZ fusion and the plasmid-borne dinB-lacZ fusion would then be that lexA protein binds more weakly to the plasmid-borne dinB operator than to the dinB operator on the chromosome. Such an alteration in  $K_d$  could be due to different degrees of superhelicity between the plasmid and the chromosome. An interesting prediction of this model is that damage-inducible genes whose operators have a weak affinity for the lexA repressor will induce in a recA430 background while those with stronger affinities will remain uninduced.

## References

- Bagg, A., C.J. Kenyon, and G.C. Walker. 1981. Proc. Natl. Acad. Sci. USA, 78:5749-5753.
- Casadaban, M.J. and S.N. Cohen. 1980. J. Bacteriol. 143:971-980.
- Chernin, L.S. and V.S. Mikoyan. 1981. Plasmid 6:119-140.
- Devoret, R. and Blanco, M. 1970. M. Molec. Genet. 107, 272-280.
- Doubleday, O.P., M.H.L. Green, and B.A. Bridges. 1977. J. Gen. Microbiol. 101:163-166.
- Elledge, S.J. and G.C. Walker. 1983. J. Mol. Biol. 163:
- George, J. M. Castellazzi, and G. Buttin. 1975. Molec. Gen. Genet. 140:309-332.
- Kato, T. and Y. Shinoura. 1977. Mol. Gen. Genet. 156:121-131.
- Kenyon, C.J. and G.C. Walker, 1980. Proc. Natl. Acad. Sci. USA 77:2819-2823.
- Kenyon, C.J., R. Brent, M. Ptashne, and G.C. Walker. 1982. J. Mol. Biol. 160:445-457.
- Langer, P.J., W.G. Shanabruch and G.C. Walker. 1981. J. Bacteriol. 145:1310-1316.
- Langer, P.J. and G.C. Walker. 1981. Mol. Gen. Genet. 182:268-272.
- Little, J.W. and D.W. Mount. 1982. Cell. 29:11-22.
- McCann, J., N.W. Spingarn, J. Kobori, and B.N. Ames. 1975. Proc. Natl. Acad. Sci. (Wash.) 72:979-983.
- Miller, J.H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Miura, A., and J. Tomezawa. 1968. Mol. Gen. Genet. 103:1-10.
- Morand, P., Blanco, M. and Devoret, R. 1977. J. Bacteriol. 131:572-582.
- Mortelmans, K.E. and Stocker, B.A.D. 1979. Mol. Gen. Genet. 167:317-328.
- Mount, D.W., K.B. Low, and S. Edmiston. 1972. J. Bacteriol. 112, 886-893.
- Mount, D.W. 1977. Proc. Natl. Acad. Sci. USA 78:300-304.
- Perry, K.L. and G.C. Walker. 1982. Nature 300:278-281.
- Roberts, J. and Roberts, C. 1981. Nature 290:422-242.
- Shanabruch, W.G. and G.C. Walker. 1980. Mol. Gen. Genet. 179:289-297.
- Steinborn, G. 1978. Mol. Gen. Genet. 165:87-93.
- Walker, G.C. 1977. Mol. Gen. Genet. 152:93-103.
- Walker, G.C. 1978. J. Bacteriol. 135:415-421.
- Walker, G.C. and P.P. Dobson. 1979. Mol. Gen. Genet. 172:17-24.
- Witkin, E. 1976. Bacteriol. Rev. 40:869-907.

## Chapter IV

### Introduction

The similarities between the umuD/C and mucA/B loci are considerable. Both loci code for two polypeptides with molecular weights of 16 and 45 kilodaltons respectively (Perry and Walker, 1982; Elledge and Walker, 1983, this thesis). Both loci appear to be organized in an operon with the smaller protein (mucA, umuD) transcribed first (Elledge and Walker, 1983, this thesis). Both loci are regulated in a negative fashion by the lexA protein in a manner consistent with the lexA gene product acting as the direct repressor of transcription (Elledge and Walker, 1983; Bagg *et al.*, 1981). Moreover, since the mucA/B genes can genetically suppress mutations in the umuD/C genes, (Walker and Dobson, 1979), it seems likely that the two sets of proteins have similar functions. Furthermore, although inclusion of pKM101 in the strain used for detecting mutations in the lacI system increases the mutagenesis rate for NCS eight fold, it does not alter the spectrum of mutagenesis (P. Foster personal communication), further supporting a functional similarity between the two proteins. This evidence strongly supports the notion that these proteins are related to one another through a common evolutionary origin.

In order to determine the exact nature of the relationship between these sets of proteins, it became desirable to define the nucleotide sequence of these genes. Having the nucleotide sequence allows us to determine a number of facts. First, it will allow us to determine the amino acid sequence of the proteins and thus

discover the extent of their evolutionary divergence. Conserved amino acid sequences may also yield important information on which areas of the protein molecule are important in a functional sense. Secondly, it allows us to decide between an operon structure for the proteins and that of a polyprotein which is processed into the 16 and 45 kilodalton proteins by proteolysis, an alternative completely consistent with the available data. Third, it allows us to investigate the likelihood that the lexA protein is the direct repressor by comparison of the operator-promoter sequence with those for other lexA protein binding sites. Finally, determination of the nucleotide sequence will facilitate manipulation of the gene for the construction of over-expressors and site-directed mutagenesis.

The work described herein was done in collaboration with Barbara Mitchell, technician par excellence. The sequencing of these genes has not been completed at the time of this thesis. However, the data thus far has yielded a number of interesting observations.

## Results

### Nucleotide sequence of umuD and the amino terminus of umuC genes and the deduced amino acid sequence.

In collaboration with Barbara Mitchell the entire nucleotide sequence of umuD and the amino terminal segment of the umuC gene was determined using the Sanger chain termination method (Sanger et al., 1977). A restriction map showing the positioning of the umuD and umuC proteins and the direction and extent of the sequenced regions is illustrated in Figure 4-1. We chose to sequence the 1 kb Bgl II fragment because it contained the operator-promoter region, the entire umuD gene and the amino terminus of the umuC gene. This was known because I cloned the 1 kb Bgl II fragment into pMC874 and generated a protein fusion as I did for mucB in chapter III. This fusion plasmid was damage-inducible for  $\beta$ -galactosidase activity, and therefore contained the operator-promoter region and was found to complement umuD mutations and therefore contained the umuD gene (data not shown).

### The regulatory region of the umuDC operon.

An autoradiogram of DNA sequencing gel of an Mp8 clone (USB22) containing most of the putative operator-promoter region and the translational start of the umuD gene is shown in Fig. 4-2. This sequence and the sequences before it comprising nucleotides 17-86 of Fig. 4-4 have several features which suggest that it is the regulatory region: 1) It is located in the area where we suspected the start of the umuD protein would be from our insertion mutagenesis studies; 2)

Figure 4-1. Restriction Map of the umuD/C locus showing the positions of peptides and the direction and extent of sequencing.

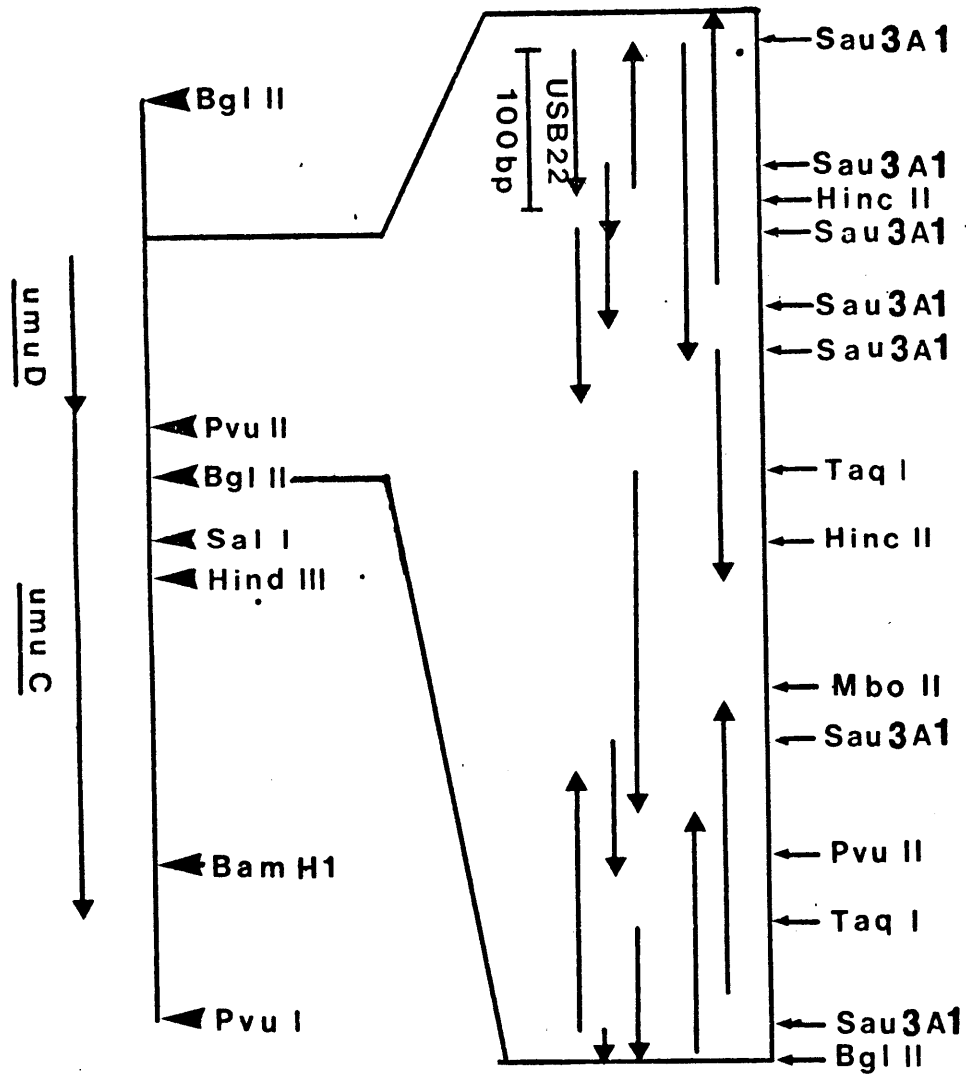
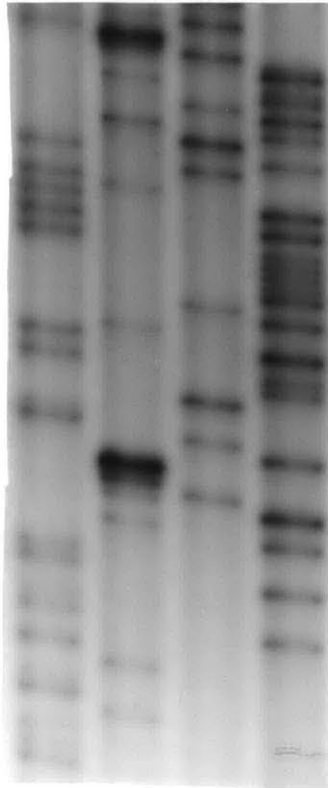


Figure 4-1



Figure 4-2. A DNA sequencing gel of a fragment containing part of the operator-promoter region and translational start of the umuD gene.

**T G C A**



ACA  
GAC  
TAC  
TGT  
ATA  
TAA  
AAA  
CAG  
TAT  
AAC  
TTC  
AGG  
CAG  
ATT  
ATT  
ATG - Met  
TTG - Leu

It contains 2 possible RNA polymerase binding sites, one comprising nucleotides 17-22 TATGAT and a weaker possibility comprising nucleotides 41-46, TACTGT; 3) It contains two tandem lexA protein binding sites separated by only 4 nucleotides; 4) It contains sequences complementary to the ribosome binding site on the 16S rRNA (Watson, 1976) (nucleotides 66-78); 5) It contains a methionine codon (ATG) directly after its ribosomal binding site at nucleotide 80 which is then followed by an open reading frame 360 nucleotides long. In fact, this region contains everything one would expect for a regulatory region with the exception of a -35 sequence for RNA polymerase binding which would be just beyond where our sequences start.

The most likely candidate for an RNA polymerase binding site is nucleotides 17-22, TATGAT, which deviates only slightly from the consensus "Pribnow box" TATAAT (Siebenlist et al., 1980). This promoter sequence is identical to the promoter sequence for tyr-tRNA (Siebenlist et al., 1980). The other possible promoter TACTGT, nucleotides 41-45, deviates significantly from the consensus sequence. This promoter is identical to the promoter for the ara BAD operon (Sienbenlist et al., 1980; Lee, 1978). That promoter is known to require two positive activators for transcription in vivo, the araC protein and CRP plus cAMP. Apparently this promoter sequence is not sufficient in and of itself to promote efficient transcription and therefore is the less likely of the two promoter candidates. It must be noted that this entire sequence may not be the actual promoter-regulatory region of the operon. There is still DNA remaining to be sequenced on the Bgl II fragment upstream of our

Figure 4-3. A. Comparison of the lexA protein binding sites of the umuD/C operon with sequences of known lexA protein binding sites.

B. Functional organization of the umuD/C regulatory region.

C. Comparison of ribosomal binding sites of the umuD and umuC genes with the 16S rRNA sequence.

Figure 4-3.

<u>umuO</u> 1	A T C T G C T G G C A A G A A C A G A C	This thesis
<u>umuO</u> 2	T A C T G T A T A T A A A A C A G T A	This thesis
<u>recA</u>	T A C T G T A T G A G C A T A C A G T A	Brent and Ptashne, 1981     "
<u>lexA</u> 1	T G C T G T A T A T A C T C A C A G C A	
<u>lexA</u> 2	A A C T G T A T A T A C A C C C A G G G	"
<u>uvrA</u>	T A C T G T A T A T T C A T T C A G G T	Sancar and Rupp, 1982
<u>uvrB</u>	A A C T G T T T T T T T A T C C A G T A	Sancar <i>et al.</i> , 1982
<u>colE1</u>	T G C T G T A T A T A A A C C A G T G	Ebina <i>et al.</i> , 1981
<u>cloDF13</u>	T A C T G T G T A T A T A T A C A G T A	van der Elzin <i>et al.</i> , 1982
<u>sfiA</u>	T A C T G T A C A T C C A T A C A G T A	Beck and Bremer, 1980
conserved	- - C T G T? - - - - - - - - C A G - -	

AGTATGATCTGCTGGCAAGAACAGACTACTGTATATAAAACAGTATA

Regulatory region of umuD/C

Pribnow                      umuO 1                      umuO 2  
box

umuD                                      MetLeu . . .  
CTTCAGGCAGATTATTATGTTG.....  
  •••••  
16S rRNA                      AUUCCUCCACUAG  
  
  MetPhe . . .  
umuC                                      GTTAAGGCGATGCGCTGATGTTT.....  
  •••••  
16S rRNA                      AUUCCUCCACUAG

Ribosomal binding site  
and translational start  
of the umuD and umuC genes

putative regulatory region. Furthermore, RNA transcription studies must be performed in vitro to define the true promoter.

The comparison between the putative lexA protein binding sequences of the umuDC operon and other known lexA protein binding sequences is shown in Figure 4-3. Sancar et al., (1982) have suggested that the rotationally symmetrical sequence CTG.....CAG, which is conserved in all damage-inducible genes operators sequenced to date, is the most important determinant of lexA protein binding. It should be noted that these sequences are separated by exactly 10 nucleotides, one turn of the helix, so that they will be adjacent to one another in space. The umuDC operon contains two such operator sequences separated by only four nucleotides. umuO1, defined by nucleotides 23-38, contains six of the seven bases conserved by the other operator sites. Besides the conserved sequences, it shares little with other operators. In fact, it resembles umuO2, defined by nucleotides 43-58, more closely than any of the other operators due to its use of As on the right half of its sequence. umuO2 shares a great deal of homology with the remaining operators, especially colE1 with which it shares 17 of the 20 bases shown here. The significance of the adjacent operator-like sequences for umuDC is not known. The lexA gene has two binding sequences before itself and binds the lexA protein cooperatively (R. Brent, Ph.D. Thesis). One difference is that the operators are separated by five bases instead of four for umuDC. Perhaps the lexA protein binds the umuDC operators cooperatively also.

Another consistent feature of these binding sites is that binding of the lexA protein to umuO1 would most certainly block binding of RNA

polymerase to the putative promoter, TATGAT, at nucleotides 17-22 and could also affect the second possible promoter site at nucleotides 41-46. Binding to umuO2 would certainly block binding to the second promoter and could also possibly affect binding to the first promoter if not just blocking its polymerization path. I would like to comment here on my use of this terminology umuO to refer to these putative operator sites. Operator sites are defined by mutation, DNA footprinting analysis and inhibition of transcription studies. None of these have been performed on these putative operators and some must be performed before they can be truly defined.

#### Open reading frames.

The nucleotide sequence for the region indicated in Fig. 4-1 is shown in Fig. 4-4 along with the deduced amino acid sequence. Starting with the methionine codon ATG at nucleotide 80, we have an open reading frame until the double stop codons at nucleotide 440. This region is enough to code for a polypeptide 120 amino acids long with an approximate molecular weight of 13,200 daltons. If there was deletion of a base between nucleotides 370 and 440 it would shift the protein into the lower reading frame and extend the open reading frame to the stop at nucleotide 485. This would yield a protein of 132 amino acids with an approximate molecular weight of 14,500 daltons. Likewise an insertion of a base between nucleotides 415 and 440 would shift one into a reading frame extending to nucleotide 498 giving a 15,000 dalton protein. The umuD protein has a molecular weight of 16,000 daltons estimated by SDS polyacrylamide gel electrophoresis

Figure 4-4. Nucleotide sequence of the umuD gene and the beginning of the umuC gene along with the deduced amino acid sequence. The suspected reading frames for the umuC and umuD proteins are underlined.



Figure 4-4

SEQUENCE FILE: FND.DAT

ASP SER LEU \*\*\* SER VAL \*\*\* SER ALA GLY LYS ASN ARG LEU LEU TYR ILE LYS THR VAL  
 ILE ALA CYS ASN GLN TYR ASP LEU LEU ALA ARG THR ASP TYR CYS ILE \*\*\* LYS GLN TYR  
 \*\*\* PRO VAL ILE SER MET ILE CYS TRP GLN GLU GLN THR THR VAL TYR LYS ASN SER ILE  
 G A T A G C C T G T A A T C A G T A T G A T C T G C T G G C A A G A A C A G A C T A C T G T A T A T A A A A C A G T A  
 10 20 30 40 50 60

\*\*\* LEU GLN ALA ASP TYR TYR VAL GLY LEU SER SER LEU ARG ILE SER GLY ASN CYS \*\*\*  
 ASN PHE ARG GLN ILE ILE MET ILE VAL TYR GLN ALA CYS GLY SER PRO GLU ILE VAL ASP  
 THR SER GLY ARG LEU LEU CYS TRP PHE ILE LYS PRO ALA ASP LEU ARG LYS LEU LEU THR  
 T A A C T T C A G G C A G A T T A T T A T G T T G G T T T A T C A A G C C T G C G G A T C T C C G G A A A T T G T T G A  
 70 80 90 100 110 120

PRO PHE PRO LEU PHE SER ASP LEU VAL GLN CYS GLY PHE PRO SER PRO ALA ALA ASP TYR  
 LEU PHE ARG TYR ILE ALA ILE SER ILE PHE SER VAL TRP ALA PHE SER PHE THR GLY SER ARG ILE THR  
 C C T T T T C C G C T A T T T A G C G A T C T T G T T C A G T G T G G C T T C C T T C A C C G G C A G C A G A T T A C  
 130 140 150 160 170 180

VAL GLU GLN ALA HIS ARG SER GLU SER THR VAL ASP GLN HIS PRO ALA ARG LEU THR SER  
 LEU ASN LYS ARG ILE ASP LEU ASN GLN LEU ILE SER ILE GLN ARG ASP LEU LEU ARG  
 \*\*\* THR SER ALA SER ILE \*\*\* ILE ASN GLN CYS \*\*\* SER ALA SER SER ALA THR TYR PHE VAL  
 G T T G A A C A A G C G C A T C G A T C T G A A T C A A C T G T T G A T C A G C A T C C A G C G C G A C T T A C T T C G  
 190 200 210 220 230 240

SER SER LYS TRP \*\*\* PHE TYR ASP \*\*\* TRP TRP ASN \*\*\* \*\*\* ARG \*\*\* PHE THR ASP CYS  
 GLN ALA SER GLY ASP SER MET ILE ASP GLY GLY ILE SER ASP GLY ASP LEU ILE ILE VAL  
 LYS GLN VAL VAL ILE LEU \*\*\* ILE MET VAL GLU LEU VAL THR VAL ILE TYR \*\*\* LEU SER  
 T C A A G C A A G T G G T G A T T C T A T G A T T G A T G G T G G A A T T A G T G A C G G T G A T T T A C T G A T T G T  
 250 260 270 280 290 300

ARG \*\*\* ARG TYR TYR ARG GLN PRO TRP \*\*\* TYR CYS HIS ARG CYS CYS \*\*\* ARG ARG VAL  
 ASP SER ALA ILE THR ALA SER HIS GLY ASP ILE VAL ILE ALA ALA VAL ASP GLY GLN PHE  
 ILE ALA LEU LEU PRO PRO ALA MET VAL ILE LEU SER SER LEU LEU LEU THR ALA SER LEU  
 C G A T A G C G C T A T T A C C G C C A G C C A T G G T G A T A T T G T C A T C G C T G C T G T T G A C G G C G A G T T  
 310 320 330 340 350 360

TYR GLY GLU LYS ILE \*\*\* ARG ASN THR PRO THR VAL GLN LEU ILE PRO \*\*\* GLN ARG VAL  
 THR VAL LYS LYS PHE SER ALA ILE ARG ARG ARG TYR SER LEU PHE HIS ASP SER ALA TYR  
 ARG \*\*\* LYS ASN LEU ALA GLN TYR ALA ASP GLY THR ALA TYR SER MET THR ALA ARG MET  
 T A C G G T G A A A A A T T T A G C G C A A T A C G C C G A C G G T A C A G C T T A T T C C A T G A C A G C G C G T A  
 370 380 390 400 410 420

CYS PRO LEU PRO SER VAL VAL LYS ILE GLN LEU ASP VAL PHE GLY VAL VAL ILE HIS VAL  
 ALA HIS TYR HIS GLN \*\*\* \*\*\* ARG TYR SER TRP MET SER LEU VAL LEU \*\*\* SER THR SER  
 PRO ILE THR ILE SER SER GLU ASP THR ALA GLY CYS LEU TRP CYS CYS ASP PRO ARG ARG  
 T G C C C A T T A C C A T C A G T A G T G A A G A T A C A G C T G G A T G T C T T T G G T G T T G T G A T C C A C G T C  
 430 440 450 460 470 480

VAL LYS ALA MET ARG \*\*\* CYS LEU PRO SER VAL MET \*\*\* THR ARG PHE MET PRO ALA VAL  
 LEU ARG ARG CYS ALA ASP VAL CYS PRO LEU \*\*\* CYS LYS ARG VAL LEU CYS GLN LEU \*\*\*  
 \*\*\* GLY ASP ALA LEU MET PHE ALA LEU CYS ASP VAL ASN ALA PHE TYR ALA SER CYS GLU  
 G T T A A G G C G A T G C G C T G A T G T T G C C C T C T G T G A T G T A A A C G C G T T T A T G C C A G C T G T G  
 490 500 510 520 530 540

ARG ARG VAL SER PRO TYR LEU TRP VAL ASN ARG TRP LEU CYS TYR ARG ILE MET THR VAL  
 ASP VAL PHE ARG PRO ILE TYR GLY \*\*\* THR GLY GLY CYS ALA ILE GLU \*\*\* \*\*\* ARG LEU  
 THR CYS PHE ALA LEU PHE MET GLY LYS PRO VAL VAL LEU SER ASN ASN ASP GLY CYS  
 A G A C G T G T T T C G C C C T A T T T A T G G G T A A A C C G G T G G T T G T G C T A T C G A A T A A T G A C G G T T  
 550 560 570 580 590 600

ALA LEU SER PRO GLU THR LEU ARG GLN ARG ARG LEU ALA LEU LYS TRP ALA ILE PRO GLY  
 ARG TYR ARG PRO LYS ARG \*\*\* GLY GLY LYS GLY ALA TRP ARG \*\*\* ASN GLY ARG SER LEU VAL  
 VAL ILE ALA ARG ASN ALA GLU ALA LYS ALA LEU GLY VAL LYS MET GLY ASP PRO TRP PHE  
 G C G T T A T C G C C C G A A A C G C T G A G G C A A A G G C G C T T G G C G T A A A A T G G G C G A T C C C T G G T  
 610 620 630 640 650 660

SER ASN LYS LYS ILE  
 GLN THR LYS ARG SER  
 LYS GLN LYS ASP LEU  
 T C A A A C A A A A G A T T T  
 670

4 14 24 34 44

(this thesis). This is within the ballpark for the molecular weights listed above. Unfortunately, much of the sequence between nucleotides 250 and 450 was done on only one strand, so the likelihood of frame-shifts is significant enough to make assignment of the exact reading frame tricky. Further evidence mentioned later suggests that the tandem double stop codons are the likely endpoint for the umuD gene.

60 nucleotides separate the end of the coding region for umuD and the beginning of the umuC coding region at nucleotide 500. The ATG codon at nucleotide 500 is preceded by a region, nucleotides 484-498, with extensive homology to the 16S rRNA. The reading frame extends for 276 nucleotides to the end of the Bgl II fragment. We anticipate the open reading frame to continue for at least 1 kilobase of DNA further into the gene to account for the 45 kilodalton umuC protein. We feel fairly confident of this sequence, in general, for several reasons. First, the entire coding region has been sequenced in both directions and often in the same direction multiple times. Secondly, the reading frame is open and ends in Asp at the Bgl II site which was predicted because of the protein fusion generated using pMC874 (Casadaban et al., 1980). The third reason will be discussed in the next section.

The fact that there are clearly two proteins coded for here allows us to lay to rest the notion of the processed polyprotein. Furthermore, since there are no promoter-like sequences in the intercistronic region combined with the evidence cited in chapter II, it appears as if the umuD and umuC genes are truly arranged in an operon.

Comparison of the umuC and mucB nucleotide  
and deduced amino acid sequences.

The nucleotide sequence of a segment of the mucAB operon was determined. This segment consisted of 280 nucleotides and contained the translational start signal of the mucB gene and an open reading frame for 198 nucleotides. This reading frame remained open to the end of the sequence thus far generated. The ATG codon at nucleotide 86 of the sequence is preceded by a stretch of DNA with homology to the 16S rRNA, nucleotides 71-84. This entire region has been sequenced in both directions, often more than once in the same direction.

We compared this DNA sequence and the deduced amino acid sequence with the corresponding region of umuD DNA sequence. The results are illustrated in Fig. 4-5. The homology of the amino acid sequence is striking. Of 61 amino acids (mucB has an extra amino acid inserted between amino acid 28 and 29 of the umuC protein) exactly 30 are conserved. Furthermore, the conservation occurs in runs. Of the first 15 amino acids, 12 are conserved. Of the next 12 amino acids only 1 is conserved. Amino acids 28-36 has 8 out of 9 conserved. Amino acids 37-56 have 9 of 19 conserved. The last five amino acids are totally unconserved.

At the nucleotide level 98 of 184 nucleotides are conserved. The true conservation may be even higher because a -1 frameshift appears to have occurred at nucleotide 152 of the muc sequence. This threw the next nine bases, which would have been perfect conservation, out of frame for comparison. An insertion event of +4 restores the

Figure 4-5. A. Comparison of the nucleotide sequence of the amino terminal end of the umuC and mucB genes.

B. Comparison of the deduced amino acid sequence of the umuC and mucB gene.

Figure 4-5

A. Nucleotide Sequence (Nucleotides 435-674 of umuC Fig. 4-4)

umuC AGTAGTGAAGATACAGCTGGATGTCTTTGGTGTGTGATCCACGTCGTTAAGGCCGATGCGCTG497  
mucB CATCCATAATGAGAGCGTTGAAATCTGGGGAGTGGTTACGCATTCCCTTATCGAGCATCCGGT 85  
 Tandem Stop Codons  
 Met Phe Ala  
umuC ATG TTT GCC CTC TGT GAT GTA AAC GCG TTT TAT GCC AGC TGT GAG ACG545  
mucB ATG TTT GCG CTG ATT GAT GTC AAT GGC ATG TAC GCC AGC TGT GAG CAG133  
umuC TGT TTC GCC CTA TTT ATG GGT AAA CCC GTG GTT GTG CTA TCG AAT AAT593  
mucB GCA TTT AGG CCA GAT CTG GCA AAC CGA GTG GCC GTT TTA TCC AAC AAT181  
 GCA - Extra Codon in mucB  
umuC GAC GGT TGC GTT ATC GCC CGA AAC GCT GAG GCA AAG GCG CTT GGC GTT641  
mucB GAC GGC AAC ATT GTG GCC CGT AAT TAC CTG GCA AAG AAA CTG GCG AAG229  
umuC AAA ATG GGC GAT CCC TGG TTC AAA CAA AAA GAT674  
mucB AAA GCG GGC CTC AAA TGG CGA TCG TAC TTC AAA262

## B. Deduced Amino Acid Sequence

umuC Met Phe Ala Leu Cys Asp Val Asn Ala Phe Tyr Ala Ser Cys Glu Thr  
mucB Met Phe Ala Leu Ile Asp Val Asn Gly Met Tyr Ala Ser Cys Glu Gln  
umuC Cys Phe Ala Leu Phe Met Gly Lys Pro Val Val Val Leu Ser Asn Asn  
mucB Ala Phe Arg Pro Asp Leu Ala Asn Arg Val Ala Val Leu Ser Asn Asn  
 Ala - Extra Amino Acid in mucB  
umuC Asp Gly Cys Val Ile Ala Arg Asn Ala Glu Ala Lys Ala Leu Gly Val  
mucB Asp Gly Asn Ile Val Ala Arg Asn Tyr Leu Ala Lys Lys Leu Ala Lys  
umuC Lys Met Gly Asp Pro Trp Phe Lys Gln Lys Asp  
mucB Lys Ala Gly Asp Lys Trp Arg Ser Tyr Phe Lys

proper amino acid reading frame at nucleotide 161. This event explains the extra amino acid in the mucB gene and increases the nucleotide conservation to 103 of 184. Of the 30 conserved amino acids, 15 had a change in the third base of the codon, one had a change in the first base and the remaining 14 were unchanged.

The nucleotide sequence of approximately eighty bases before the respective translational start codons of mucB and umuC has no apparent homology. One interesting fact is that approximately 55 bases before the translational start of these genes is a tandem pair of termination codons. These codons were the suspected termination codons for the umuD protein. The fact that the bases before the coding regions for umuC and mucB are totally divergent yet at approximately the correct position retain a tandem pair of stop codons suggests that these termination codons are of some functional significance to the sequence. One last fact, the protein fusion between mucB and  $\beta$ -galactosidase predicted an Asp inframe at the Bgl II site of the mucB gene. The Bgl II site is at nucleotide 147 and sure enough, there is the Asp confirming our reading frame.

### Discussion

The sequence of the umuD and part of the umuC genes have yielded a great deal of information. The sequence we have designated as the promoter-regulatory region is most probably correct. It has every feature needed to explain the properties of the regulation and production of the umuC and umuD proteins. The tandem lexA binding sequences suggest a finer-tuned regulation than was previously suspected. DNA footprinting analysis, transcription inhibition studies and run-off transcription analysis are certainly in order here to define the functionality of the suggested promoter(s) and operator(s) sites.

This study has shown conclusively that the umuD and umuC proteins are made as separate proteins and constitute an operon. Furthermore, it suggests that the umuD protein consists of 120 amino acid residues and is terminated by tandem termination codons. The authenticity of the reading frame can be checked by labeling the protein with various hot amino acids in maxicells, purifying the protein off of gels and subjecting it to proteolysis or chemical cleavage. The size of the labeled fragments should be predictable from the sequence data. Another check for the proper reading frame will be a comparison with the mucA sequence when it is eventually defined.

By far the most exciting discovery of this study is the apparent common evolutionary origin of the umuC and mucB proteins. These genes are 50% conserved on an amino acid and nucleotide level. This virtually guarantees a common biochemical function. The stretches of conserved amino acids will surely provide important information

concerning functionally important parts of the molecules. The significance of the conservation is brought out more clearly when compared to the noncoding bases directly preceding the genes. These sequences have totally diverged (29% homology) since their evolutionary separation probably due to the lack of selective pressure for a function. In other words, if a sequence can change, it will, unless it has a good reason (a needed function) not to change.

The degree of divergence of the nucleotides sequence suggests that the separation of these genes occurred a long time ago. This degree of divergence may be misleading. Since the functions of these genes promote UV and chemical mutagenesis and raise the level of spontaneous mutagenesis (and raise it still further when SOS is induced), they each probably accelerated their own divergence. This degree of divergence does explain my inability to observe a signal in Southern hybridization experiments, assuming, of course, that the remaining sequences will have diverged to approximately the same degree as the sequenced regions.

Another genetic observation concerns the relationship of umuC to umuD and mucA to mucB. If these proteins each performed a separate biochemical function, then one would expect that mucA could perform umuD's function and therefore could replace it in relation to umuC. However, I have found that mucA<sup>+</sup> mucB<sup>-</sup> plasmids fail to complement in umuD<sup>-</sup> umuC<sup>+</sup> strains (data not shown). This suggests that the umuC and umuD proteins interact and that mucA and mucB interact but that they have each co-evolved to be specific for one another. It will be interesting to test the mucA<sup>-</sup> mucB<sup>+</sup>, umuD<sup>+</sup> umuC<sup>-</sup> complementation when the strains become available. Furthermore, this interaction could be tested in vitro using radiolabeled proteins.



My final comment concerns the evolutionary significance of these genes. Plasmids often confer a number of selective advantages to their hosts. The muc genes confer a protection function which enable the hosts to survive insults to the DNA. It is tempting to speculate that these genes confer another selective advantage on their hosts, namely, the acceleration of genetic diversity under stress. In other words, evolution may select not only for the adapted, but also for the ability to adapt, the ability to evolve. Furthermore, certain selective pressures, such as UV-irradiation, amino acid starvation and thymine starvation, which are known to induce SOS, may actually accelerate the accumulation of genetic diversity because induction of SOS increases the rate of spontaneous mutagenesis and, hence, aids adaptation. This idea is well within the framework of Darwinian evolution because the cells are still evolving randomly with no a priori knowledge of what is they are adapting towards. It is, however, somewhat at variance with the notion generated by the experiments of Luria and Delbruck which is that the selection process selects for mutations that previously existed within a given population at the time of selection. In the presence of genes like mucA/B and umuD/C, certain selection processes do generate mutations. However, this would not have been discovered through a fluctuation analysis even if the mutational assay had been reversion of an amino acid auxotrophy instead of phage resistance because jackpots would occur anyway.

## References

- Bagg, A., Kenyon, C.J. and Walker, G.C. (1981) Proc. Natl. Acad. Sci. USA 78, 5749-5753.
- Beck, E. and Bremer, E. (1980) Nuc. Acids Res. 13, 3011-3024.
- Brent, R. and Ptashne, M. (1981) Proc. Natl. Acad. Sci. USA 78, 4204-4208.
- Casadaban, M.J., Chou, J. and Cohen, S.N. (1980) J. Bacteriol. 143, 971-980.
- Ebina, Y., Kishi, F., Miki, T., Kagamuyama, H., Nakazawa, T. and Nakazawa, A. (1981) Gene 15, 119-126.
- Elledge, S.J. and Walker, G.C. (1983) J. Molec. Biol. 163, in press.
- Lee, N. (1978) in The Operon, eds. Miller, J. and Reznikoff, W.S. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. p. 389.
- Perry, K.L. and Walker, G.C. (1982) Nature 300, 278-281.
- Sancar, A. and Rupp, W.D. (1982) Nature, in press.
- Sancar, G.B., Sancar, A., Little, J.W. and Rupp, W.D. (1982) Cell 28, 523-530.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Siebenlist, U., Simpson, R.B. and Gilbert, W. (1980) Cell 20, 269-281.
- van der Elzen, P.J.M., Maat, J., Walters, H.B., Veltkamp, E. and Nykamp, H.J.J. (1982) Nuc. Acids Res. 10, 1913-1928.
- Walker, G.C. and Dobsen, P. (1979) Mol. Gen. Genet. 172, 17-24.
- Watson, J. (1976) in The Molecular Biology of the Gene, W.A. Benjamin, Inc., Reading, MA.

## Chapter V

### Materials and Methods

#### Bacterial strains, media and chemicals.

The bacterial strains employed in this study are described in Table 5-1. Rich medium employed was LB (Miller, 1972). The minimal base medium used was M9 (Miller, 1972) supplemented with 0.2% glucose and amino acids to a final concentration of 50 µg/ml. When included, kanamycin sulfate was used at 30 µg/ml, ampicillin at 25 µg/ml, chloramphenicol at 30 µg/ml, spectinomycin at 100 µg/ml, thymine at 50 µg/ml, adenine at 100 µg/ml and 5-bromo-4-chloro-3-indoyl-β-D-galactoside (Xgal) at 40 µg/ml. Spectinomycin was a gift from the Upjohn Co. Adenine, thymine, amino acids, ortho-nitrophenylgalactoside and all other antibiotics were purchased from Sigma Chemical Co. Xgal was obtained from Bachem Inc., and all restriction enzymes and DNA ligase were purchased from New England Biolabs and used under the suppliers' suggested conditions.

#### Plasmid Construction

All plasmid DNA used for cloning were purified by CsCl-ethidium bromide buoyant density centrifugation. Ligations were typically performed at 10 µg/ml of both vector and insert DNAs with T4 DNA ligase (New England Biolabs) at 4°C for 10 hours. DNA fragments from partial digests were purified from 1% low melting agarose gels (BRL). Transformations were performed as in Cohen et al. (1972). All plasmids used are listed in Table 5-2.

Table 5-1. Bacterial Strains

Strain	Genotype	Reference or Source
AB1157	F <sup>-</sup> , <u>thr-1</u> , <u>leu-6</u> , <u>proA2</u> , <u>his-4</u> , <u>thi-1</u> , <u>argE3</u> , <u>lacY1</u> , <u>galK2</u> , <u>ara-14</u> , <u>xyl-5</u> , <u>mtl-1</u> , <u>tsx-33</u> , <u>rpsL31</u> , <u>supE44</u> , $\lambda^-$	A.J. Clark
DM49	as AB1157, but <u>lexA3</u> (Ind <sup>-</sup> ) (1972)	D. Mount <u>et al.</u>
JC2926	as AB1157, but <u>recA13</u>	A.J. Clark
JC10521	as AB1157, but <u>recA430</u> ( <u>lexB30</u> ), <u>thy</u>	A.J. Clark
GC3217	F <sup>-</sup> , <u>thr-1</u> , <u>leu-6</u> , <u>proA2</u> , <u>his-4</u> , <u>argE3</u> , <u>galK2</u> , <u>rpsL31</u> , <u>ilv(ts)</u> , <u>recA441</u> ( <u>tif-1</u> ), <u>sulA11</u> ( <u>sfiA11</u> )	George <u>et al.</u> (1975)
DML180	as GC3217 but <u>arg</u> <sup>+</sup> <u>lexA3</u> (Ind <sup>-</sup> )	D. Mount (1977)
DM1415	as DML180 but <u>recA56</u> , <u>lexA51</u> ( <u>spr-51</u> ), <u>his</u> <sup>+</sup>	D. Mount
KM1190	as DM1415 but <u>recA</u> <sup>+</sup> , <u>srlC5</u>	T. Roberts
GW2010	as KM1190 but <u>recA430</u> ( <u>lexB30</u> ), <u>srl::Tn10</u>	C. Kenyon
GW1000	as GC3217 but <u>pro</u> <sup>+</sup> , <u>lac</u> $\Delta$ (U169)	Kenyon and Walker (1980)
GW1030	as GW1000 but <u>dinB::Mu d</u> (Ap, <u>lac</u> )	This Paper

Strain	Genotype	Reference or Source
GW1039	as GW1030 but <u>recA430</u> ( <u>lexB30</u> ), <u>srl::Tn10</u>	This paper
RB901	F <sup>-</sup> , <u>thr-1</u> , <u>leu-6</u> , <u>lacY?</u> , <u>rpsL31</u> , <u>mtlA</u> , $\Delta$ ( <u>recA srl</u> )21, <u>lexA51</u> (Def), <u>lexA3</u> (Ind <sup>-</sup> ), <u>sulA1</u>	R. Brent
TK610	as AB117 but <u>arg</u> <sup>+</sup> , <u>uvrA6</u> , <u>ilv325</u> , <u>umuC36</u>	T. Kato, (Kato and Shimoura, 1977)
GW1003	as AB1157 but <u>zcf289::Tn5</u>	C. Kenyon
GW1103	as GW1000 but <u>uvrA6</u> , <u>umuC121::Mu d</u> (Ap, <u>lac</u> )	(Bagg <u>et al.</u> 1981)
GW558	as AB1157 but <u>purB15</u>	
GW2100	as AB1157 but <u>umuC122::Tn5</u>	
MM294A	<u>pro82</u> , <u>thi1</u> , <u>endA1</u> <u>hsdR17</u> , <u>supE44</u>	K. Backman
GW3200	as AB1157 but <u>umuC44</u>	
GW3199	as AB1157 but <u>umuC77</u>	
SY822	$\Delta$ ( <u>lacZ</u> )445, <u>lacI3</u> , $\Delta$ ( <u>recA-srl</u> )304, <u>his-4</u> , <u>argE3</u> , <u>rpsL31</u> , <u>xyl-5</u> <u>mtl-1</u>	R. Isberg

Table 5-2. Plasmids

Plasmids	Relevant Genotype	Source
pBR322	Ap <sup>R</sup> Tc <sup>R</sup>	
pSC101	Tc <sup>R</sup>	K. Backman
pDPT427	Sp <sup>R</sup> Cm <sup>R</sup>	D. Taylor (Sninsky <u>et al.</u> 1981)
pL10	Ap <sup>R</sup> , <u>gpt</u> <sup>+</sup> derived from pBR322	R. Mulligan
pRB160	Ap <sup>R</sup> , <u>lexA</u> <sup>+</sup>	R. Brent
pMC874	Km <sup>R</sup> , <u>lacZ</u> <sup>-</sup> protein fusion vector	C. Kenyon (Casaban and Cohen, 1980)
pMC903	Km <sup>R</sup> , <u>lacZ</u> operon fusion vector	"
pKB280	Tc <sup>R</sup> , <u>cI</u> <sup>+</sup>	(Backman and Ptashne, 1978)
pKB354	Ap <sup>R</sup> Tc <sup>R</sup>	K. Backman
pKM101	Ap <sup>R</sup> , <u>mucA</u> <sup>+</sup> <u>mucB</u> <sup>+</sup>	
pGW270	Ap <sup>R</sup> , <u>mucA</u> <sup>+</sup> <u>mucB</u> <sup>+</sup> deletion derivative of pKM101	(Langer <u>et al.</u> 1980)
pGW510	Ap <sup>R</sup> , <u>dinB-lacZ</u> <sup>+</sup>	(Kenyon <u>et al.</u> 1982)
pGW1700	Tc <sup>R</sup> , <u>mucA</u> <sup>+</sup> , <u>mucB</u> <sup>+</sup> derivative of pKB354	(Perry and Walker, 1982)
pSE100	Ap <sup>R</sup> Km <sup>R</sup> derivative of pBR322	This thesis
pSE101	Km <sup>R</sup> derivative of pSC101	" "
pSE110	Km <sup>R</sup> , <u>umuC</u> <sup>+</sup> , <u>umuD</u> <sup>+</sup> derivative of pSE101	" "
pSE111	as pSE110 but containing approx- imately the same insert in the opposite orientation	" "
pSE114	Km <sup>R</sup> <u>umuC</u> <sup>+</sup> <u>umuD</u> <sup>+</sup>	" "
pSE115	Km <sup>R</sup> , <u>umuC</u> <sup>+</sup> <u>umuD</u> <sup>+</sup>	" "
pSE116	Km <sup>R</sup> , <u>umuC</u> <sup>+</sup> <u>umuD</u> <sup>+</sup> contains smallest insert	" "

Plasmids	Relevant Genotype	Source
pSE117	Ap <sup>R</sup> Km <sup>R</sup> <u>umuC</u> <sup>+</sup> <u>umuD</u> <sup>+</sup> derivative of pL10 (pBR322)	This thesis
pSE126	as pSE110 but <u>umuC101::Tn1000</u>	" "
pSE128	as pSE110 but <u>umuC105::Tn1000</u>	" "
pSE129	as pSE110 but <u>umuC106::Tn1000</u>	" "
pSE130	as pSE110 but <u>umuC107::Tn1000</u>	" "
pSE131	as pSE110 but <u>umuC109::Tn1000</u>	" "
pSE132	as pSE110 but <u>umuD111::Tn1000</u>	" "
pSE133	as pSE110 but <u>umuC112::Tn1000</u>	" "
pSE137	Sp <sup>R</sup> derivative of pSE116	" "
pSE140	as pSE110 but $\Phi(\text{umuD}'\text{-lacZ}^+)$ 140	" "
pSE151	Sp <sup>R</sup> Tc <sup>R</sup> derivative of pBR322	" "
pSE152	Sp <sup>R</sup> , <u>lexA</u> <sup>+</sup> derivative of pBR160	" "
pSE200	Km <sup>R</sup> , <u>mucA</u> <sup>+</sup> , $\Phi(\text{mucB}'\text{-lacZ}'^+)$ 200 derivative of pMC874	" "
pSE210	Cm <sup>R</sup> derivative of pGW510 which is <u>lacZ</u> <sup>-</sup>	" "

pSE100 was constructed by cloning BamHI-digested GW2100 DNA into BamHI-cleaved pBR322 by selecting  $\text{Ap}^{\text{R}}\text{Km}^{\text{R}}$ . pSE101 was constructed by cloning the HindIII-BamHI fragment containing the Tn5 neomycin phosphotransferase gene from pSE100 into HindIII-BamHI cleaved pSC101 (Cohen and Chang, 1977) and selecting  $\text{Km}^{\text{R}}$  while screening for  $\text{Ap}^{\text{S}}\text{Tc}^{\text{S}}$ . pSE110-113 were constructed by cloning 6-9 kb Sau3AI-generated partials of  $\lambda_{\text{SE14}}$  into BamHI-cleaved pSE101, selecting for  $\text{Km}^{\text{R}}$  and screening for the ability to complement the umuC36 allele of TK610. pSE114 and pSE115 were constructed in a similar manner using 2-5 kb Sau3AI partials of  $\lambda_{\text{SE14}}$ . pSE116 was constructed by replacing the HindIII fragment of pSE114 which contains the neomycin phosphotransferase gene ( $\text{Km}^{\text{R}}$ ) with the HindIII fragment of pSE115 which also contains the neomycin phosphotransferase gene. pSE137 was constructed by replacing the EcoRI-PstI fragment of pSE116 with the EcoRI-PstI fragment of pDPT427 (Sninsky et al., 1981) which contains the gene coding for spectinomycin resistance. Spectinomycin-resistant clones were selected and screened for the loss of  $\text{Km}^{\text{R}}$  and the retention of the ability to complement the umuC36 mutation of TK610. pSE140 was constructed by cloning the purified lac operon-containing BamHI fragment from pMC903 (Casadaban et al., 1980) into pSE110 DNA made linear by a very limited digest with Sau3AI followed by gel purification of full length linears. The ligation mixture was transformed into GW1000 (tif,  $\Delta\text{lac}$ ) (Kenyon and Walker, 1980) and selected on Km-Xgal (Miller, 1972) plates.  $\text{Km}^{\text{R}} \text{Lac}^+$  colonies were then assayed for the inducibility of  $\beta$ -galactosidase activity (Miller, 1972) in minimal media at  $42^{\circ}\text{C}$ . Plasmids showing 10 fold induction or better were studied further. pSE152 was constructed by replacing the



EcoRI-PstI fragment of pRB160 (a clone of the lexA<sup>+</sup> gene) (Brent and Ptashne, 1980) with the EcoRI-PstI fragment of pDPT427. Sp<sup>R</sup> clones were screened for Ap<sup>R</sup> and UV sensitivity. pSE151 was constructed by replacing the EcoRI-PstI fragment of pBR322 with the same EcoRI-PstI fragment of pDPT427.

pSE117 was constructed by cloning the EcoRI-KpnI digested pSE115 DNA into EcoRI-KpnI digested pL10 and transforming MM294A selecting for Km<sup>R</sup>Ap<sup>R</sup>. DNA from Km<sup>R</sup>Ap<sup>R</sup> colonies were analyzed for the proper EcoRI-KpnI fragments. A derivative with the desired fragments was chosen for further study and designated pSE117.

pSE200 was constructed by cloning BglIII-digested pGW270 DNA into Bam HI-cleaved pMC874 and transforming into GW1000 selecting for Km<sup>R</sup> on plates containing Xgal. Km<sup>R</sup> blue colonies were screened for Ap<sup>S</sup> to eliminate derivatives containing the BglIII fragment of pGW270 carrying the bla gene. The remaining derivatives were screened for the ability to induce  $\beta$ -galactosidase synthesis in response to treatment with 15 J/m<sup>2</sup> of UV-irradiation. Plasmid DNA's from derivatives showing induction were analyzed for insertions. All plasmids screened contained the 2.3 kb BglIII fragment of pGW270 in the orientation shown in Fig. 1. One derivative was chosen for further study and was designated pSE200.

pSE210 was constructed by cloning the HindIII-PstI fragment of pDPT427 containing the chloramphenicol resistance gene into the HindIII-PstI fragment of pGW510 which contained the pBR322 origin of replication and the operator-promoter region of the dinB::Mu d(Ap, lac) fusion but which lacked the lacZ gene. Cm<sup>R</sup> transformants of

GW1000 were isolated on plates containing Xgal and chloramphenicol. Cm<sup>R</sup> white colonies were chosen and their plasmid DNAs were isolated and screened for the presence of the 2 HindIII-PstI fragments. A plasmid which contained the desired fragments was designated pSE210.

#### DNA Preparation

Isolation of bacteriophage DNA was as reported by Shih and Weinberg (1982) with the exception that the phage were grown on AB1157. Chromosomal DNA was prepared by resuspending 200 mls of cells in 7.5 mls of 50 mM Tris pH8, 100 mM EDTA and 20% Sucrose. 1.8 mls of 20 mg/ml lysozyme in 0.25 M Tris pH8 was added and incubated for 20 minutes on ice. Then 20 µl of 10 mg/ml RNase A, 1 ml of 2 mg/ml of proteinase K in 0.25 M Tris pH8 and 10 mls of 1% Sarkosyl in 75 mM EDTA pH8 were added and the solution was incubated at 37°C for 6 hours. This solution was adjusted to a density of 1.71 g/ml with CsCl and centrifuged overnight in a Beckman L5-50 ultracentrifuge in a V-Ti50 rotor at 42,000 rpm. DNA was isolated by puncturing the tube and collecting viscous material. This was dialysed against 10 mM Tris, 1 mM EDTA pH8, and used directly for Southern blotting analysis and cloning. Plasmid DNA used for cloning was isolated by the procedure of Timmis et al. (1978) with the modification of Taylor and Cohen (1979). Mini-preps of plasmids for restriction analysis were performed according to Holmes and Quigley (1981).

### Electrophoresis of DNA and Southern Blotting

Approximately 1 µg of each DNA sample to be analyzed was digested in 50 µl using restriction enzymes purchased from New England Biolabs under conditions suggested by the supplier. Samples were electrophoresed on horizontal (16 cm x 20 cm, 0.8%) agarose gels in E buffer, pH 7.8 (Sharp et al., 1973) at 40 V for 14 hours. HindIII digested λ DNA was used as a molecular weight standard.

Gels were stained in 1 µg/ml ethidium bromide, photographed, and UV irradiated for 5 minutes with a Ultraviolet Products Inc. transilluminator in order to nick the DNA and facilitate the transfer of higher molecular weight restriction fragments. The DNA in the gels was then denatured and blotted onto nitrocellulose (Schleicher and Schuell BA85) according to the method of Southern (1975).

### Preparation of <sup>32</sup>P-labeled DNA

Specific restriction fragments of cloned DNAs were isolated from agarose gels using low melting agarose obtained from BRL. These fragments were labeled by nick translation (Rigby et al., 1977), with the modifications of Chung et al. (1981). Labeled DNA was denatured by boiling for 10 minutes and rapidly cooled in an ice water bath. Dextran Blue was added to the DNA and this was run through a Sephadex G-75 column pre-equilibrated with 0.1 M NaCl, 10 mM Tris pH 7.5. The dye containing fractions were collected and used directly for hybridizations.

### Hybridizations

Filters for  $\lambda$  plaque hybridization were prepared according to Benton and Davis (1977). The hybridization to these filters was performed in the same manner used for Southern blotting. Filters were presoaked for 10 hours at 37°C in hybridization buffer (50% formamide, 4xSSC, 0.02% BSA, 0.02% polyvinylpyrrolidone, 0.02% Ficoll, 0.1% SDS, 25 mM EDTA, 20 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.0). This solution was replaced by fresh hybridization buffer with 10<sup>5</sup>-10<sup>6</sup> cpm/ml of labeled probe, sealed in a Sears Seal'N'Boil bag and incubated at 37°C with gentle shaking for 8-12 hrs. Filters were then rinsed with 2xSSC and placed in 1 liter of 2xSSC, 0.05% SDS at 55°C-65°C for 3-4 hours. Filters were rinsed with 2xSSC again, dried, and exposed to Kodak XR-5 film using a Cronex intensifying screen.

### UV Mutagenesis and Survival

The complementation of the umuC<sup>-</sup> phenotype was assayed in several ways. Complementation by the Charon 28 clone of umuC was assayed by growing GW2100(pKB280) in  $\lambda$ ym medium to 2x10<sup>8</sup> cells/ml, harvesting the cells, resuspending in 0.85% saline, 10 mM MgSO<sub>4</sub> and adsorbing the recombinant  $\lambda$  bacteriophage at an m.o.i. of 2.5 for 15 min at 37°C. Cells were then placed in glass petri dishes and irradiated with 15 W G.E. germicidal lamp at a UV fluence of 1 J/m<sup>2</sup>/sec for varying lengths of time. Mutagenesis and survival was measured by plating the cells on supplemented M9/glucose plates in 2.5 mls of F top agar containing

a limiting amount of arginine (0.05 mM). Plates were incubated at 37°C for 3 days before counting. UV survival and mutagenesis curves for plasmid complementation were performed in the same manner as above except that cells were grown in supplemented M9/glucose medium instead of  $\lambda$ ym medium.

Screening of potential plasmid subclones for their ability to complement umuC mutants was initially accomplished by patching colonies in 2 cm<sup>2</sup> patch on supplemented M9-glucose plates containing a limiting amount of histidine (0.05 mM). The plates were UV irradiated with 2 J/m<sup>2</sup> for uvrA<sup>-</sup> strains and 15 J/m<sup>2</sup> for uvrA<sup>+</sup> strains and patches which gave 25 or more revertants were scored as umuC<sup>+</sup>, while umuC<sup>-</sup> colonies rarely showed more than 1 revertant.

#### Maxicells and Electrophoresis of Proteins

Labeling of plasmid proteins by the maxicell procedure was performed essentially as described by Sancar et al. (1981) with the following modifications. The strains bearing plasmids were RB901 ( $\Delta$ recA spr51) and SY822 ( $\Delta$ recA) and these were irradiated with 30 J/m<sup>2</sup>. After the overnight incubation 5 ml of these cells were resuspended in 1 ml of sulfate-free Hershey medium and labeled with 20  $\mu$ Ci of <sup>35</sup>S-methionine for 15 mins before harvesting. Cells were resuspended in 100 $\lambda$  of lysis buffer (10% glycerol, 5% B-mercapto-ethanol, 3% SDS, .001% Bromophenol Blue, and .0625M Tris HCl pH 6.8) and treated at 90°C for 5 mins to denature the proteins and lyse the cells.

Polyacrylamide gradient gels (10-20%) containing SDS were made using the Laemmli buffer system; the conditions for the gel are as described by Walker and Dobson (1979). Gels were stained with Coomassie Blue R250 and treated with PPO for fluorography (Bonner and Laskey 1974) before drying and exposure to film.

#### Tn1000 Mutagenesis

Tn1000 insertions in pSE110 were isolated by the method of Sancar and Rupp (1979). The donor, MM294A(F'128)(pSE110) ( $Km^R$ ) and the recipient, TK610 (umuC36,  $Sm^R$ ) were grown to  $1 \times 10^8$  cells/ml in LB, mixed in equal volumes and allowed to mate for 1 hr at  $37^\circ C$ . Plasmids in which Tn1000 insertions had occurred were screened for their ability to complement umuC36 mutation of TK610.

#### $\beta$ -Galactosidase assays.

Cells were grown overnight in LB medium containing the appropriate drugs. These were diluted one hundred fold into minimal media supplemented with 1% casamino acids and 0.2% glucose and grown to a  $3 \times 10^7$  cell/ml. Cells were placed in glass petri dishes and UV-irradiated with  $15 \text{ J/m}^2$  of UV light from a 15W G.E. germicidal lamp at a UV fluence of  $1 \text{ J/m}^2/\text{sec}$ . For induction experiments involving recA441 (tif-1) containing strains, cells were placed at  $42^\circ C$  and adenine was added to  $100 \text{ }\mu\text{g/ml}$ .  $\beta$ -galactosidase activity was determined by the method of Miller (1972).

### DNA sequencing

DNA was sequenced using the chain termination method described by Sanger et al., (1977). Single-stranded DNA was generated using M13 mp8 or mp9 as described by Messing et al., (1981) and as suggested by Bethesda Research Laboratories. DNA sequence was analyzed on an Apple II computer using the Stanford GENET and SEQ programs (Brutlag et al., 1982).

## References

- Backman, K. and Ptashne, M. (1978). Cell 13, 65-71.
- Bagg, A., Kenyon, C.J. and Walker, G.C. (1981). Proc. Natl. Acad. Sci. USA 78, 5749-5753.
- Benton, W.D. and Davis, R.W. (1977). Science 196, 180-182.
- Bonner, W.M. and Laskey, R.A. (1974). Eur. J. Biochem. 46, 83-88.
- Brent, R. and Ptashne, M. (1980). Proc. Nat. Acad. Sci. USA 77, 1932-1936.
- Brutlag, D.L., Clayton, J., Friedland, P. and Kedes, L. (1982). Nuc. Acids Res. 10, 279-294.
- Casadaban, M.J., Chou, J. and Cohen, S.N. (1979). J. Bacteriol. 143, 971-980.
- Chung, S., Landfear, S.M., Blumberg, D.D., Cohen, N.S. and Lodish, H.F. (1981). Cell 24, 785-797.
- Cohen, S.N. and Chang, A.C.Y. (1977). J. Bacteriol. 132, 734-737.
- Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972). Proc. Nat. Acad. Sci. U.S.A. 69, 2110-2114.
- George, J., Castellazzi, M. and Buttin, G. (1975). Mol. Gen. Genet. 140, 309-332.
- Holmes, D.S. and Quigley, M. (1981). Anal. Biochem. 114, 193-197.
- Jorgensen, R., Rothstein, S.J. and Reznikoff, W.S. (1979). Molec. Gen. Genet. 177, 65-72.
- Kato, T. and Shinoura, Y. (1977). Gen. Genet. 156, 121-131.
- Kenyon, C.J., Brent, R., Ptashne, M. and Walker, G.C. (1982). J. Mol. Biol. 160, 445-457.
- Kenyon, C. and Walker, G. (1980). Proc. Nat. Acad. Sci. USA 77, 2819-2823.
- Langer, P.J., Shanabruch, W.G. and Walker, G.C. (1981). J. Bacteriol. 145, 1310-1316.
- Messing, J., Crea, R. and Seeburg, P.H. (1981). Nuc. Acids Res. 9, 309-321.
- Miller, J.H. (1972) in "Experiments in Molecular Genetics". Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Mount, D.W., Low, K.B. and Edmiston, S. (1972). J. Bacteriol. 112, 886-893.
- Mount, D.W. (1977). Proc. Nat. Acad. Sci. USA 74, 300-304.
- Perry, K.L. and Walker, G.C. Nature, 300, 278-281.
- Rigby, P.W.J., Deckman, M., Rhodes, C. and Berg, P. (1977). J. Mol. Biol. 113, 237-254.
- Sancar, A. and Rupp, W.D. (1979). Biochem. Biophys. Res. Commun. 90, 123-129.
- Sancar, A., Wharton, R.P., Seltzer, S., Kacinski, B.M., Clarke, N.D., and Rupp, W.D. (1981). J. Mol. Biol. 148, 45-62.
- Sanger, F., Niklen, S. and Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA 74, 5463-5467.
- Sharp, P.A., Sugden, B. and Sambrook, J. (1973). Biochemistry 12, 3055-3063.
- Shih, C. and Weinberg, R.A. (1982). Cell 29, 161-169.
- Sninsky, J.J., Uhlin, B.E., Gustafsson, P. and Cohen, S.N. (1981). Gene 16, 275-286.
- Southern, E.M. (1975). J. Mol. Biol. 98, 503-517.
- Taylor, D. and Cohen, S. (1979). J. Bacteriol. 137, 92-104.
- Timmis, K., Cabello, F. and Cohen, S.N. (1978). Mol. Gen. Genet. 162, 121-137.
- Walker, G.C. and Dobson, P.P. (1979). Molec. Gen. Genet. 172, 17-24.